



# Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

## CONSULTING EDITORS

JOSEPH B. ADELSON  
*University of Michigan*

GORDON W. ALLPORT  
*Harvard University*

NORMAN H. ANDERSON  
*University of California, Los Angeles*

LEONARD BERKOWITZ  
*University of Wisconsin*

JACK W. BREHM  
*Duke University*

URIE BRONFENBRENNER  
*Cornell University*

DONALD R. BROWN  
*Bryn Mawr College*

DONALD T. CAMPBELL  
*Northwestern University*

LAUNOR F. CARTER  
*System Development Corporation*

DWIGHT W. CHAPMAN, JR.  
*Vassar College*

MORTON DEUTSCH  
*Bell Telephone Laboratories*

SEYMOUR FESHBACH  
*University of Pennsylvania*

DONALD W. FISKE  
*University of Chicago*

NATHANIEL L. GAGE  
*University of Illinois*

NICHOLAS HOBBS  
*Peabody College for Teachers*

MURRAY HORWITZ  
*New York University*

IRVING L. JANIS  
*Yale University*

HAROLD H. KELLEY  
*University of California, Los Angeles*

HERBERT C. KELMAN  
*Harvard University*

DANIEL J. LEVINSON  
*Harvard Medical School*

DAVID C. McCLELLAND  
*Harvard University*

ROBERT B. MACLEOD  
*Cornell University*

HAROLD L. RAUSH  
*National Institute of Mental Health*

NEVITT SANFORD  
*Stanford University*

IRVING SARNOFF  
*Western Reserve University*

WILLIAM A. SCOTT  
*University of Colorado*

JOSEPH VEROFF  
*University of Michigan*

JUSTIN L. WEISS  
*Massachusetts Mental Health Center*

Acting Production Manager: HELEN ORR

Editorial Assistant: JUDY SMITH

## VOLUME 64

## 1962

PUBLISHED BY  
**THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.**

Prince and Lemon Sts., Lancaster, Pa.  
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Copyright 1962 by the American Psychological Association, Inc.

# CONTENTS OF VOLUME 64

## ARTICLES, CRITIQUE AND NOTES

	PAGE
ALLPORT, FLOYD H. <b>A Structuronomic Conception of Behavior: Individual and Collective I. Structural Theory and the Master Problem of Social Psychology</b> .....	3
ARNOLD, JACK. See CHANCE, ERIKA.	
BANKS, JAMES H., JR. See MILLER, ROBERT E.	
BARBER, THEODORE XENOPHON, & GLASS, LOUIS BENJAMIN. <b>Significant Factors in Hypnotic Behavior</b> ..	222
BAXTER, JAMES C., & BECKER, JOSEPH. <b>Anxiety and Avoidance Behavior in Schizophrenics in Response to Parental Figures</b> .....	432
BECKER, JOSEPH. See BAXTER, JAMES C.	
BECKER, SELWYN W. See HOWELLS, LLOYD T.	
BERKOWITZ, LEONARD, & GREEN, JAMES A. <b>The Stimulus Qualities of the Scapegoat</b> .....	293
BLAKE, ROBERT R., & MOUTON, JANE SRYGLEY. <b>Overevaluation of Own Group's Product in Intergroup Competition</b> .....	237
BOSLEY, JOHN J. See DE SOTO, CLINTON B.	
BRAMEL, DANA. <b>A Dissonance Theory Approach to Defensive Projection</b> .....	121
BROCK, TIMOTHY C. <b>Cognitive Restructuring and Attitude Change</b> .....	264
BURGESS, MICHAEL. See HOKANSON, JACK E.	
BURNSTEIN, EUGENE, & MCRAE, ADIE V. <b>Some Effects of Shared Threat and Prejudice in Racially Mixed Groups</b> .....	257
CAPRA, PAUL C., & DITTES, JAMES E. <b>Birth Order as a Selective Factor among Volunteer Subjects</b> .....	302
CASTORE, GEORGE F. <b>Number of Verbal Interrelationships as a Determinant of Group Size</b> .....	456
CERBUS, GEORGE, & NICHOLS, ROBERT C. <b>Personality Correlates of Picture Preferences</b> .....	75
CHANCE, ERIKA, ARNOLD, JACK, & TYRRELL, SYBIL. <b>Communality and Stability of Meaning in Clinical Case Description</b> .....	389
DECHARMS, RICHARD, & MOELLER, GERALD H. <b>Values Expressed in American Children's Readers: 1800-1950</b> .....	136
CLARIDGE, G. See EYSENCK, H. J.	
COHEN, JACOB, & STRUENING, E. L. <b>Opinions about Mental Illness in the Personnel of Two Large Mental Hospitals</b> .....	349
COHEN, MELVIN. See KATZ, IRWIN.	
COTTRELL, NICKOLAS B. See STOTLAND, EZRA.	
CURTIS, THEODORE. See ZANDER, ALVIN.	
DAVIDS, ANTHONY, KIDDER, CATHERINE, & REICH, MELVYN. <b>Time Orientation in Male and Female Juvenile Delinquents</b> .....	239
DE SOTO, CLINTON B., & BOSLEY, JOHN J. <b>The Cognitive Structure of a Social Structure</b> .....	303
DEWOLFE, ALAN S. <b>The Effect of Affective Tone on the Verbal Behavior of Process and Reactive Schizophrenics</b> .....	450
DITTES, JAMES E. See CAPRA, PAUL C.	
EDWARDS, ALLEN L., & WALKER, JERALD N. <b>Relationship between Probability of Item Endorsement and Social Desirability Scale Value for High and Low Groups on Edwards' SD Scale</b> .....	458
EPSTEIN, SEYMOUR, & FENZ, WALTER D. <b>Theory and Experiment on the Measurement of Approach-Avoidance Conflict</b> .....	97
EPSTEIN, SEYMOUR, & LEVITT, HERBERT. <b>The Influence of Hunger on the Learning and Recall of Food Related Words</b> .....	130
EYSENCK, H. J., & CLARIDGE, G. <b>The Position of Hysteries and Dysthyniics in a Two-Dimensional Framework of Personality Description</b> .....	46
FEDER, WALDA. See GOLDSTEIN, MICHAEL J.	
FELD, SHEILA. See VEROFF, JOSEPH.	
FELDSTEIN, STANLEY. <b>The Relationship of Interpersonal Involvement and Affectiveness of Content to the Verbal Communication of Schizophrenic Patients</b> .....	39
FENZ, WALTER D. See EPSTEIN, SEYMOUR.	
FERGUSON, EVA DREIKURS. <b>Ego Involvement: A Critical Examination of Some Methodological Issues</b> ..	407
GLADIS, MICHAEL, & WISCHNER, GEORGE J. <b>Schizophrenic and Normal Response Patterns to "Aversive" and "Neutral" Associations in Two Paired-Associate Paradigms</b> .....	249
GLASS, LOUIS BENJAMIN. See BARBER, THEODORE XENOPHON.	
GOLDSTEIN, MICHAEL J., & HIMMELFARB, SAMUEL. <b>The Effects of Providing Knowledge of Results upon the Perceptual Defense Effect</b> .....	143



	PAGE
GOLDSTEIN, MICHAEL J., HIMMELFARB, SAMUEL, & FEDER, WALDA. A Further Study of the Relationship between Response Bias and Perceptual Defense.....	56
GREEN, JAMES A. See BERKOWITZ, LEONARD.	
GURIN, GERALD. See VEROFF, JOSEPH.	
GYNTH, MALCOLM D. Crime and Psychopathology.....	378
HANDELMAN, NORMAN S. See LACHMANN, FRANK M.	
HANFORD, DANA B. See SINNETT, E. ROBERT.	
HARBURG, ERNEST. See HOFFMAN, L. RICHARD.	
HILLMER, MAX L., JR. See STOTLAND, EZRA.	
HIMMELFARB, SAMUEL. See GOLDSTEIN, MICHAEL J.	
HOFFMAN, L. RICHARD, HARBURG, ERNEST, & MAIER, NORMAN R. F. Differences and Disagreement as Factors in Creative Group Problem Solving.....	206
HOKANSON, JACK E., & BURGESS, MICHAEL. The Effects of Three Types of Aggression on Vascular Processes.....	446
HOLLAND, BERT. See SPENCE, DONALD P.	
HOWELLS, LLOYD T., & BECKER, SELWYN W. Seating Arrangement and Leadership Emergence.....	148
KATZ, DANIEL. Editorial.....	1
KATZ, IRWIN, & COHEN, MELVIN. The Effects of Training Negroes upon Cooperative Problem Solving in Biracial Teams.....	319
KIDDER, CATHERINE. See DAVIDS, ANTHONY.	
KUETHE, JAMES L. Social Schemas.....	31
LACHMANN, FRANK M., LAPKIN, BENJAMIN, & HANDELMAN, NORMAN S. The Recall of Dreams: Its Relation to Repression and Cognitive Control.....	160
LANG, PETER J., & LOUTO, KENNETH. Mediation and Associative Facilitation in Neurotic, Psychotic, and Normal Subjects.....	113
LAPKIN, BENJAMIN. See LACHMANN, FRANK M.	
LEE, JOAN C., & TUCKER, R. BENNETT. An Investigation of Clinical Judgment: A Study in Method.....	272
LEVENTHAL, HOWARD, & PERLOE, SIDNEY I. A Relationship between Self-Esteem and Persuasibility.....	385
LEVITT, HERBERT. See EPSTEIN, SEYMOUR.	
LONGENECKER, E. D. Perceptual Recognition as a Function of Anxiety, Motivation, and the Testing Situation.....	215
LORR, MAURICE, MCNAIR, DOUGLAS M., MICHAUX, WILLIAM W., & RASKIN, ALLEN. Frequency of Treatment and Change in Psychotherapy.....	281
LUOTO, KENNETH. See LANG, PETER J.	
MCCORD, JOAN, MCCORD, WILLIAM, & THURBER, EMILY. Some Effects of Paternal Absence on Male Children.....	361
MCCORD, WILLIAM. See MCCORD, JOAN.	
McGEE, RICHARD K. The Relationship between Response Style and Personality Variables: I. The Measurement of Response Acquiescence.....	229
McGUIRE, WILLIAM J. Persistence of the Resistance to Persuasion Induced by Various Types of Prior Belief Defenses.....	241
MCNAIR, DOUGLAS M. See LORR, MAURICE.	
MCRAB, ADIE V. See BURNSTEIN, EUGENE.	
MAHLER, IRWIN. Yeasayers and Naysayers: A Validating Study.....	317
MAIER, NORMAN R. F. See HOFFMAN, L. RICHARD.	
METZNER, RALPH. See MISCHEL, WALTER.	
MICHAUX, WILLIAM W. See LORR, MAURICE.	
MILLER, ROBERT E., BANKS, JAMES H., JR., & OGAWA, NOBUYA. Communication of Affect in "Cooperative Conditioning" of Rhesus Monkeys.....	343
MISCHEL, WALTER, & METZNER, RALPH. Preference for Delayed Reward as a Function of Age, Intelligence, and Length of Delay Interval.....	425
MOELLER, GERALD H. See DECHARMS, RICHARD.	
MOUTON, JANE SRYGLEY. See BLAKE, ROBERT R.	
NICHOLS, ROBERT C. See CERBUS, GEORGE.	
OGAWA, NOBUYA. See MILLER, ROBERT E.	
PAUL, COLEMAN. Effect of Distracting Words on a Size Estimation Task.....	79
PERLOE, SIDNEY I. See LEVENTHAL, HOWARD.	
RASKIN, ALLEN. See LORR, MAURICE.	
RAU, LUCY. See WINDER, C. L.	
REECE, MICHAEL M., & WHITMAN, ROBERT N. Expressive Movements, Warmth, and Verbal Reinforcement.....	234
REICH, MELVYN. See DAVIDS, ANTHONY.	



	PAGE
SAMPSON, EDWARD E. Birth Order, Need Achievement, and Conformity .....	155
SANDLER, JACK. The Effect of Negative Verbal Cues upon Verbal Behavior.....	312
SAXTON, GEORGE H. Spontaneous Fantasy as a Resource of High Grade Retardates for Coping with a Failure-Stress Frustration .....	81
SHANNON, DONALD T. Clinical Patterns of Defense as Revealed in Visual Recognition Thresholds.....	370
SILVERMAN, DORIS K. See SILVERMAN, LLOYD H.	
SILVERMAN, LLOYD H., & SILVERMAN, DORIS K. Ego Impairment in Schizophrenia as Reflected in the Object Sorting Test.....	381
SINNETT, E. ROBERT, & HANFORD, DANA B. The Effects of Patients' Relationships with Peers and Staff on Their Psychiatric Treatment Programs.....	151
SMITH, PHILIP A. A Comparison of Three Sets of Rotated Factor Analytic Solutions of Self-Concept Data	326
SNOEK, J. DIEDRICK. Some Effects of Rejection upon Attraction to a Group.....	175
SPENCE, DONALD P., & HOLLAND, BERT. The Restricting Effects of Awareness: A Paradox and an Explanation .....	163
STOTLAND, EZRA, & COTTRELL, NICKOLAS B. Similarity of Performance as Influenced by Interaction, Self-Esteem, and Birth Order.....	183
STOTLAND, EZRA, & HILLMER, MAX L., JR. Identification, Authoritarian Defensiveness, and Self-Esteem	334
STRUENING, E. L. See COHEN, JACOB.	
THURBER, EMILY. See MCCORD, JOAN.	
TUCKER, R. BENNETT. See LEE, JOAN C.	
TYRRELL, SYBIL. See CHANCE, ERIKA.	
UPSHAW, HARRY S. Own Attitude as an Anchor in Equal-Appearing Intervals.....	85
VEROFF, JOSEPH, FELD, SHEILA, & GURIN, GERALD. Dimensions of Subjective Adjustment.....	192
WALKER, JERALD N. See EDWARDS, ALLEN L.	
WEIDENFELLER, EDWARD W., & ZIMNY, GEORGE H. Effects of Music upon GSR of Depressives and Schizophrenics .....	307
WHITMAN, ROBERT N. See REECE, MICHAEL M.	
WINDER, C. L., & RAU, LUCY. Parental Attitudes Associated with Social Deviance in Preadolescent Boys .....	418
WISCHNER, GEORGE J. See GLADIS, MICHAEL.	
WORELL, LEONARD. Response to Conflict as Determined by Prior Exposure to Conflict.....	438
ZANDER, ALVIN, & CURTIS, THEODORE. Effects of Social Power on Aspiration Setting and Striving.....	63
ZIMNY, GEORGE H. See WEIDENFELLER, EDWARD W.	



# Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

## CONSULTING EDITORS

JOSEPH B. ADELSON  
*University of Michigan*

GORDON W. ALLPORT  
*Harvard University*

NORMAN H. ANDERSON  
*University of California, Los Angeles*

LEONARD BERKOWITZ  
*University of Wisconsin*

JACK W. BREHM  
*Duke University*

URIE BRONFENBRENNER  
*Cornell University*

DONALD R. BROWN  
*Bryn Mawr College*

DONALD T. CAMPBELL  
*Northwestern University*

LAUNOR F. CARTER  
*System Development Corporation*

DWIGHT W. CHAPMAN, JR.  
*Vassar College*

MORTON DEUTSCH  
*Bell Telephone Laboratories*

SEYMOUR FESHBACH  
*University of Pennsylvania*

DONALD W. FISKE  
*University of Chicago*

N. L. GAGE  
*University of Illinois*

NICHOLAS HOBBS  
*Peabody College for Teachers*

MURRAY HORWITZ  
*New York University*

IRVING L. JANIS  
*Yale University*

HAROLD H. KELLEY  
*University of California, Los Angeles*

HERBERT C. KELMAN  
*Harvard University*

DANIEL J. LEVINSON  
*Harvard Medical School*

DAVID C. MCCLELLAND  
*Harvard University*

ROBERT B. MACLEOD  
*Cornell University*

HAROLD L. RAUSH  
*National Institute of Mental Health*

NEVITT SANFORD  
*Stanford University*

IRVING SARNOFF  
*Western Reserve University*

WILLIAM A. SCOTT  
*University of Colorado*

JOSEPH VEROFF  
*University of Michigan*

JUSTIN L. WEISS  
*Massachusetts Mental Health Center*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.



The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is \$10.00, or \$20.00 annually (foreign, \$20.50). Single issues are \$2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ  
Department of Psychology  
University of Michigan  
Ann Arbor, Michigan

---

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR  
Acting Production Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 Sixteenth Street, N.W.  
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
Prince and Lemon Sts., Lancaster, Pa.  
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Lancaster, Pa., and at additional mailing offices

Copyright © 1962 by the American Psychological Association, Inc



## EDITORIAL

Psychological journals are in the dilemma of serving two functions: the older purpose of a forum of broad communication and discussion, and the ever-growing function of a research repository for experimental findings. The assumption in the first purpose is that the journal has readers who regard it as their medium for the exchange of ideas. Two forces, however, are destroying this journal function. Increasing specialization of knowledge means that whole areas of research investigation in a broad field are of only peripheral concern to research workers absorbed in their own subarea. And research output is growing at a faster rate than journal space, so research reporting becomes increasingly cryptic and technical and of ready comprehension only to the initiates of a given ingroup. The logical development of this trend is the demise of journals as journals and the publication of an annotated bibliography with microfilm cards containing the essence of the experiment available on request to interested researchers. Problems of information storage and retrieval are in fact receiving careful study, and the APA is currently conducting its own investigation of the communication and storage of scientific information under a National Science Foundation grant.

In a scientific area in which there is a standard nomenclature and in which the operational definitions of concepts are widely accepted, the movement toward codification, storage, and retrieval with use of electronic devices is irresistible. In the area of abnormal psychology, personality dynamics, and social functions covered by this journal, we are affected by the general trend, but less strongly and very unevenly. Some subfields covered by the *Journal of Abnormal and Social Psychology* lend themselves fairly readily to codified reporting and to standardized procedures. Other areas, however, are more appropriately served by a journal of discussion which permits the development of the rationale for an attack upon a problem, interpretative consideration of findings, and theoretical speculation. Recognition of the varied content and approaches in our domain means, then, a set of multiple criteria. It would be premature to have as our goal research articles which are one step away from retrieval through microcards. On the other hand, the page allocation from the APA based upon cost figures sets immediate limits upon the amount of discussion in the journal.

Are there any changes in policy which should be inaugurated with a new administration of

*JASP* which might improve the character of the compromises necessary to serve the dual functions of concise research reporting and broad communication needs? It would seem the better part of wisdom to follow the policies of the past editor whose stewardship of the journal during the past 6 years has done so much to raise the standards of research reporting and whose prodigious efforts and superb judgment have made the *JASP* such a prestigious journal. Nevertheless, new problems arise and old problems are often interpreted in a different frame of reference by a new editor so that modifications in policy are almost inevitable. Moreover, it may be possible within the framework already established to clarify objectives and effect some changes in practice. For example, though the policy of the journal has been one of accepting theoretical and research articles, the image of the journal is that of a periodical devoted to research reporting. This image means that very few theoretical articles are in fact submitted. To break this closed circuit and to change the image we need to publish more theoretical articles even if they do not represent tight, formal systems. With the space pressures we face, something will have to be sacrificed in the process. The sacrifice may have to be in articles of a purely methodological sort. Methodology may have to appear merely as part of the procedure section of a substantive article. Similarly, the development of measures of personality assessment may have to be published elsewhere. And we may also have to give little space to what Don Hebb calls "porcupine research"—the reporting of minute variations of some older finding which contributes nothing of a theoretical nature. The availability of a technology does not justify the ringing of all the changes it makes possible. And we will also have to scrutinize studies which either in experimental and sample design or in clarity of findings are in the nature of pilot pieces of research and which should be followed by further experimentation before publication. This is, of course, not new policy but a reminder to the present editors of important old policy.

In general the *JASP* accepts papers of three types which make contributions to abnormal and social psychology: theoretical articles which are systematic or formal in nature, reports of research, and case studies. Research reports can deal either with empirical studies or with hypothesis testing experiments. Empirical studies should go either wide or deep and have implica-



tions for psychological theory. Hypothesis testing studies can be less extensive in their coverage but should indicate where the hypotheses came from and how they are derived.

An established policy which will be maintained is the use of two basic criteria in the evaluation of research papers. A research report may qualify for publication because it represents a dependable contribution to existing knowledge through the soundness of its design and the clarity and reliability of its findings. Or it may be acceptable, even though not definitive in its results, because it opens up a new and promising area of research. The person who blazes a trail should not necessarily be asked to make it a paved highway.

Since the *JASP* does not have to face immediate conversion to annotated titles and microcards, our editorial policy will not move toward standardization of writing style in an effort to reach desirable standards of clear and rigorous presentation. A journal in which all articles might have been written by the same author may lose as much in reader interest as it gains in ready reporting. It is possible that if we were to read more and write less, we would not slow down the advance of knowledge.

Authors at the present time are under such constraints of brevity and of stylized communication that they often lose their readers early in their presentation. They are urged, therefore, to consider as their audience a broader public of psychologists than their own immediate disciples. One minor step which they should take is to reduce the use of private code letters for designating concepts, groups, and experimental manipulations.

The *JASP* has a special problem in that many of the studies it reports are concerned with differences between various groups characterized by differing psychological attributes. It is curious that the logic of sampling has often been re-

stricted in psychological research to large scale surveys. Many other investigations which attempt to generalize about differences between populations also need to deal with the sampling problem in their research design. The ignoring of the problem is due in part to the persistence of the layman's notion that any handful of people or observations is truly representative; in part to the practical difficulties of obtaining either representative or matched samples, and in part to the theoretical difficulties created by the need to define the universes from which samples are desired (e.g., schizophrenics, character disorders, etc.). It would impede progress if we were to insist upon rigorous sampling in many areas in which the limitations upon the investigator preclude such procedures. Nevertheless, the investigator should be aware of the problem and should take steps to avoid some of the pitfalls of selective bias. Moreover, he should be careful that his generalization of findings does not run ahead of his research design.

With this issue, the *JASP* moves into monthly publication. The page allotment for the year will not be increased, but the monthly form of publication makes it easier for the APA Central Office and the printer to cut down on the production lag which has plagued the journal for the past year.

An editorial enterprise of the magnitude of the *JASP* must call upon the cooperation of many scholars in the field both as members of the editorial board and as special reviewers. The incoming editor has drawn heavily upon such help and is already indebted to a long list of consultants and reviewers. Finally, mention should be made of the patience of the authors who generally would prefer to leave the verdict of the merit of their articles to history rather than to the heavy blue pencil and rejection bias of a fallible editor.

DANIEL KATZ



# A STRUCTURONOMIC CONCEPTION OF BEHAVIOR: INDIVIDUAL AND COLLECTIVE

## I. STRUCTURAL THEORY AND THE MASTER PROBLEM OF SOCIAL PSYCHOLOGY<sup>1</sup>

FLOYD H. ALLPORT<sup>2</sup>

*Syracuse University*

It is probably unwise to regard any broad problem of science as a subject that can be irrevocably closed. Though the past is strewn with errors we have learned to discard, there is always the chance that somewhere along the way certain significant but controversial issues have, like sleeping dogs, been merely allowed to lie. Sooner or later these issues may rise again to challenge us. There is one such question which, though it has been bypassed and largely forgotten in the main trend of social thinking, has never been laid aside so far as the writer is concerned.

### GROUP VERSUS INDIVIDUAL AGENCY: SEMANTIC AND LOGICAL CONFUSIONS

The story began for the writer on the day when Hugo Münsterberg assigned to him as a problem for his doctoral dissertation the investigation of the possible "influence of the group upon the mental processes of the individual." The exact details of this investigation, which were carried out in due course, are not important in the present context; but that such an influence was unmistakably shown, that it was measurable, that the results have been confirmed by others, and finally that a "group" influence has been taken for

granted as the basis for so many social theories, are facts which have posed for the writer a searching question. In what terms should the "social influence" actually be understood and explained?

There had been much writing upon the subject in an earlier day that seemed nebulous. The older theorists were often inclined toward a heavyhanded social determinism. In some instances they hypostatized a plane of "collective reality" quite distinct from the plane of the individuals. One recalls the superindividual doctrine of Durkheim, and at a later date the cultural determinism and "superorganic" theories of Kroeber and other anthropologists. There was also Le Bon's crowd consciousness and the "crowd mind," and Ross' "planes and currents" in a society. From a somewhat different point of view we had the old "social organism" metaphor, elaborated in modern times by Spencer and others, as well as the "group mind" concept of McDougall. Though strangely endowed with human qualities, the "collective" reality in most of these theories was nevertheless supposed to be above the individuals and somehow to control or direct their actions.

With respect to everyday parlance that is still current even among social scientists, it has seemed to the writer that there is also something unfortunate in the practice of employing psychological (that is, individually oriented) concepts as personifications applying to groups. If, for example, it is said that a "group convenes," "thinks," "feels," "decides," "achieves solidarity," "becomes organized," "legislates," or "adjourns," either the expression is a tautology, a borrowing for the group of a term that has meaning only at the level of the individual organism, or else it refers to some kind of "concerted doing"

<sup>1</sup> Editor's Note. As originally written, this critique of systems of social psychology included a structurally oriented examination, also, of network theories, open systems, and theory of social roles. A second article, continuing this series in the journal, will take a different course and will proceed more directly with the exposition and testing of the writer's theory of event-structure, the collective aspects of which are foreshadowed in the present analysis.

<sup>2</sup> Grateful acknowledgement is made to the administrative officers of Syracuse University for financial assistance in the preparation of this article.

The author is now Professor of Social and Political Psychology Emeritus of Syracuse University.

performed by an alleged agent for which we can find no unique referent. An *individual* can be said to "think" or "feel"; but to say that a group does these things has no ascertainable meaning beyond saying that so many individuals do them. On the other hand, an individual can walk into a room with others and sit down with them, he can arrange to do certain things with them, and he may answer "Aye" when his name is called in order to express his agreement. But the individual does not "convene," "become organized," "legislate," or "achieve solidarity." Such expressions have meaning only as designations of action at the "group" level. But here we run into a difficulty. No matter how closely we look, we can never actually experience any group agent who is performing such actions. If we point to, or try to touch, the group, we are doing nothing that can be distinguished from pointing to or touching individuals. Hence we must realize either that the description we give of the act makes it appropriate for individuals (but only figurative for a group), or else that it is appropriate for groups, but whenever we allude to any particular group that does it we are using a term (and concept) that has no unambiguously denotable referent. In this latter case the group as such, if it exists at all, must be always an unencounterable "something" of which we have no direct or sensible evidence. We have "constructed" a singular, nondenotational, group *concept* out of a plurum of experiences of elements (individuals) each of whom, separately, can be actually encountered.

This sort of usage seemed to the writer to reflect a vagueness in the social sciences with respect to important matters that contrasted strangely with ideals of precision in other disciplines. Did it represent an imprecision not only in terms, but in the very meanings the terms were supposed to express? Could it be that when one tries to put into language his thoughts and observations about what a group actually "does" (as, for example, when some cooperative act is performed), he really does not know *what* he means because he has never seriously tried to think the matter out?

There are no doubt many persons, even in the social disciplines, to whom such matters are not of vital consequence—who will say

that so long as we get useful results out of "group action" there is no real need to define it. This, of course, is a legitimate position, since the choice of methods and problems is always likely to be in part a matter of individual values. Such a view, however, is different from the evaluation here implied, and if the reader should happen to hold strongly to the opinion expressed above, he would probably find the present article pointless and might do well to lay it aside. For the conviction here expressed is that in using language to refer to "social" action it is important to know what we mean.

A social scientist or a social psychologist who has been prone to the usage we are criticizing might reply that he certainly knows what *he* means. He uses such terms, he may say, only as convenient "fictions." If, for example, we had to observe (and go through a statement of) all the hundreds of acts of individuals which are involved in a piece of legislation, we should never get anywhere in the social scene. We can sum it all up with great economy by saying "Congress" legislates. Is this not, he might ask, the way all scientists work? The physicist cannot observe atoms directly. He sets up the concept of an atom as a useful "construct" inferable from "results" or from other forms of evidence. In the same way the molar concepts of learning in psychology are hypothetical constructs, used not as clear-cut realities, but only to aid our thinking.

In answer to this argument the writer would raise the following question. Is our knowledge of what the phrase "Congress legislates" means gained from observing *acts of Congress* in the same (empirical) way that the physicist's knowledge of what interionic attraction means is gained by observing (even though indirectly) the *behavior of atoms*, or in the same way that the psychologist's description of the learning of the rat is gained by watching the *rat* as it runs the maze? Or, is the word "Congress" in the phrase "Congress legislates" a term whose meaning is established only on the basis of evidence from *other sorts* of things, such, for example, as acts that *individuals* are seen to do? To make the cases comparable would we not have to assume that the physicist had had no tech-



niques whatever that had revealed to him (through a chain of observable physical happenings) encounters with "atoms," that he had first known, by direct or indirect encounter, only electrons, protons, and neutrons, and their behavior, and that he had *then* proceeded to construct the atom *conceptually* out of these particles? These assumptions, which might have justified the social scientist's analogy, are all contrary to the actual sequence of discoveries in the physical science. The psychologist, similarly, does not construct the learning phenomenon, in the first instance, out of neurons and their properties; he begins by the molar observation of "the rat" and its macroscopically observed movements. In both of these cases observation of what happens properly begins at a level which is at the time available to denotational (that is, encountering) procedures.

The word "Congress," in contradistinction to these instances, has *actual* denotational reference only at a step that is lower down in the natural hierarchy than the level to which it is presumed to refer. Its referent is encounterable only at the level of *individuals* and their actions. Ought we not, therefore, to observe *the acts of the individuals* and base our generalizations upon *them* (that is, upon the experiences actually obtained from the phenomenon through our senses) rather than attempt to describe what *Congress* is said to do? If we were to set ourselves to study Congress and "its" actions, would we not be starting our investigation by choosing as its object something that cannot check us to tell us whether, in experiencing it, we are actually experiencing *something*, and that cannot tell us where this "something" leaves off and some other object begins? How, with such criteria omitted at the very start, could it be claimed that the social sciences are empirical? <sup>3</sup>

<sup>3</sup> A social scientist might reply that it is only the accident of our "smaller" size relative to the group that renders the group, unlike the objects of the physicist and psychologist, incapable of encounter as such. If we had much duller senses, or were far out in space looking down, the "group" would then appear "as an object" and would be available denotationally for experiments, just as the "atom" now is. Though the logic of this argument seems sound, it really betrays its originator. For the group as *then* experienced would certainly *not* be the group

#### THE INDIVIDUALISTIC POSITION: ATTACK ON "GROUP FALLACIES"—THE SOCIOLOGISTS' REPLY

It was near the beginning of such reflections that the writer decided to make a clean sweep, and concluded that although the individual behaves differently in a "group" than when alone, this fact offered no evidence of the existence of a "social" as distinct from an "individual" entity. The explanation of the so-called social influence *lay only in the psychology of the individual as he operated in situations with others*. The writer (1924b) presented this viewpoint systematically in his *Social Psychology* and followed it with numerous theoretical and experimental studies extending the same thought to such topics as conformity, customs, culture, public opinion, political organization, and collective action generally (Allport, 1924a, 1924b, 1927a, 1927b, 1933, 1934a, 1934b, 1937, 1939, 1940, 1942; Allport & Hartman, 1924, 1931). His *Institutional Behavior* (1933) documented, in addition, the psychological and ethical consequences of the indiscriminate use of group and institutional fictions in collective or social living. This program of "debunking" socio-deterministic conceptions made up in vigor all that it lacked in tact.

On the scenes that followed, the writer would like to draw the curtain. To say the least, it was scarcely a love feast. Some stoutly denied that sociologists were to any serious extent guilty of such a confusion of language as the writer had charged. The usual reaction seemed to be that it was only a matter of figurative or terminological convenience, and that of course they did not mean any of these terms literally. The implication was that in using group-agency terminology they knew exactly what they were talking about, or else that no one should require a phenomenon so familiar as a "group" or "group action" to be further defined; but if they *were* pressed for a statement, they could readily translate

phenomenon as the social scientist now conceives and studies it. The hope of a useful theoretical or practical approach to the *social* scene would seem to lie rather in our ability to construct some sort of model that will show both the individual and the collective levels *simultaneously*—that will describe, in other words, how the latter *is made up of the former*.

these matters into language that would satisfy the most meticulous. A more deliberate answer was that although some sociologists might have erred through the "group fallacy" in borrowing the psychologists' terms to apply to collective entities, their error was only skin deep. It was due to their slowness in developing a proper vocabulary for group reference. In the course of time they would be able to devise their own terminology and the problem would disappear.

Still others, like the psychologist Asch (1952b), maintained that although it was true that "group facts" have their foundation in "individual facts," and group consciousness, group purposes, and group goals are found only in individuals, "social fields" must be recognized as existing cognitively *within the individual*. Even if the group is not a separate entity, the fact that the individual often conceives it as such makes all the difference in his behavior.<sup>4</sup> To sociologists in general it did not seem that the writer's program of ascribing all causality to individual motivation could provide much enlightenment concerning social control, social institutions, culture, or similar topics which were their own primary interests.

#### DEFINING THE ISSUE

The writer readily concedes that in 1924 he did not realize the full implications of the problem. He was worrying only about the *personification* involved in the "group mind" and about the usage that he called the "institutional fallacy," when he should have been worrying also about groups and institutions themselves. That is to say, he should have been trying to examine and learn *what sort* of reality the referents of these words might actually have. Regardless of the adequacy of their own conceptual equipment, in doubting that the writer's treatment of social psychology as a purely "individual" discipline could solve their problem sociologists certainly had a point. One can, of course, discount as il-

<sup>4</sup> This position, though relevant to the problem of the group and individual in the broader sense, was not a definitive answer to the question of whether or in what sense, a "group" really exists. Nor did it render any less appropriate the attack upon terminological laxity and confusion which the writer was making.

logical the familiar argument that because individuals behave in a crowd as they would not behave alone, the crowd is therefore a mental entity that "embraces" or "descends upon" its members. But even if we got rid of the *crowd mind*, the problem of describing the differential of crowd-like behavior, that is, the phenomenon of a crowd as such, would remain. Granting that the individual is the only referent to which the terms of psychology can be applied, there remains the problem of stating in unambiguous terms, and not merely for this collectivity or that, but for collectivities in general, just what is occurring in the phenomena to which *we allude* when we say that the group "has solidarity," "is organized," "decides," or "controls its members."

Though such expressions cannot lead us to an unambiguous referent for the term "group," nor can they themselves be useful starting points for an empirically oriented investigation, one cannot for a moment deny that certain sets of happenings referred to by these terms do occur. Such happenings are important in social science and in daily life. It is also clear that an individual determinist who, in describing the individual as the agent, isolates him from acting "collectively" with others, will be equally unable to cope with them.

But unfortunately these realizations at once place us in a dilemma. Granted that there are realities that can be seen only by looking at individuals as such, and also realities that can be experienced or demonstrated only when individuals are taken together, how can these realities, in both cases, be stated in terms that are explicit, clear, and general? If we cannot be anthropomorphic in referring to groups for fear of running into a group tautology or fallacy, how then *can* we describe them? If thoroughgoing individual causality and group determinisms are either inadequate or illogical, what then is left? What sort of agency *can* be invoked? Or, if these two antithetical determinants are to be combined, who can provide the magic formula by which it is to be done?

Nothing that has been said above should be construed as a lessening of the *need* for a solution of this problem. We are merely pointing out the difficulties. It is not conceded that



social scientists' intent in the use of group fictions has been trivial, or that such fictions were used "merely" for convenience, or that the remedy could have lain in a mere change of terminology. Suppose we had gotten rid of the group fallacy by refusing to assign to the group any of the terms of description that belong to individuals. Our logic would then have become impeccable; but of what would our discourse have consisted? What *are* social phenomena if they are not the behaviors of individuals? Or, if social realities *are* entirely composed of individual actions, is there some way of describing and aggregating the latter, not before realized, that will hold the key to the statement of both realities simultaneously and without personification, tautology, or hypostatized agency? This is the difficult question by which sociologists and social psychologists inevitably *would have been faced* had they given up their use of fictions; and it is a question which, in the writer's judgment, they were not prepared to answer. It was not merely that they did not have the right words for their concepts; *the concepts themselves were lacking*.

And this issue still remains as unfinished business. In spite of advances in many areas of social psychology and the social sciences, we are still faced, if we view the situation candidly, with the realization that our theories of social causality, and of the relation of action at the "societal" level to what we recognize as the acts of individuals, are far from adequate either in denotation or explanatory significance. Again, the writer has no quarrel with those who do not think a methodological housecleaning in these respects is necessary, those who believe that it is possible to achieve useful results through individuals in groups without facing such apparently unanswerable questions. He would only state his opinion that unless we are to be content with good works alone, unless we are willing to accept the bankruptcy of the social sciences as empirical disciplines that are also denotational, we must try to discover some more satisfactory paradigm for "the group and the individual." Some way must be found to describe in general terms that layout of conditions surrounding and involving individuals which we have called "the

group," and to formulate, in the precise yet *universal* manner of science, what actually goes on in the situation we call "collective" action. In this broader sense the problem of the individual and the group is really the "master problem" of social psychology.

The present article and one to follow in a later issue are intended as a contribution to the solution of this problem. In this article an attempt will be made to trace some of the major developments that have occurred with respect to it in contemporary systems of social psychology. Every system, of course, has had to make some kind of terms with it, if only by implication. The reader's indulgence is asked for the recital of theories that may be already familiar, a procedure rendered unavoidable by the need of showing the specific ways in which their authors have dealt with the problem.<sup>5</sup> To these synopses of the theories there will be added in most cases a critical commentary by the present writer, prepared mainly with its following orientation. In two previous publications the writer (1954, 1955) has offered preliminary conceptualizations of the "patterned" or "articulated" aspect of phenomena, a model which he has hypothesized as being general and therefore as underlying integrated behavior both in the lone individual and in collective aggregates. He now believes that the formulations of the group-individual relationship that are to be found in the schools of modern social psychology will, if analyzed, reveal in latent form some of the outlines of this same structural pattern; or, if they are lacking, then by restating of the theory in such a manner as to take them into account its validity and usefulness can be enhanced. It would follow that if, in the working through of these systems, a further endeavor is made to reveal such "structural" strands and to reinterpret or extend the ideas which they involve, additional light might be thrown upon the prob-

<sup>5</sup> For additional statements of the group-individual problem and its modern setting see Doob (1952) and the very thoughtful chapter on "An Introduction to Group Theory" in Asch (1952b). (See also Allport, 1942, 1961.) The latter reference contains further suggestions on the logic of the problem of collective reality. We shall not here attempt any statement of the extensive *earlier* literature.

lem in hand. Through his critique of the theories and certain other added sections the writer has endeavored to do just this.

PHILOSOPHICAL APPROACHES—COGNITIVE  
VIEWPOINTS—ATTEMPT AT A  
GESTALT SYNTHESIS

The impetus toward group reification at the time the writer began his work, though not productive of formal models, had already acquired a philosophical underpinning. There was developing a background of superorganic theory in cultural anthropology, as well as a doctrine of "emergent evolution" in biology. "Supersummation" and "wholistic properties" were added by gestalt psychology. All these ideas, if not already at the social level, could be readily extrapolated to it. As time went on, the phenomenological tendencies of the configurational psychologists had to be reckoned with. Though the writer had collected the "social introspections" of his subjects in his experiments on "group influence," there were those who felt, as Asch did, that the comparative neglect of the way the individual was *perceiving* the particular situation he was in had stood in the way of a full investigation of the group phenomenon. The "distortion effects" elicited in group experiments by Asch (1952a) could be cited as evidence of the "pressure" of the judgments of the group upon the cognitions of the individual. In such phenomena the manner in which the subject is perceiving the reactions of those about him is all-important.

The gestalt position, in conformity with Koffka (1935) and as stated more recently by Asch (1952b), can be summarized as follows: *The concepts both of the individual and the group are necessary for a full explanation of social and individual behavior, and from a psychological standpoint the two go hand in hand. Groups and institutions have their existence as indispensable constructs in the mental life of individuals.* The doctrine that there is a "whole property" that is not to be found in the parts was said to apply no less to collective than to individual aggregates. A group process "is neither a sum of individual activities nor a fact added to the activities of individuals" (Asch, 1952b, p. 252). There was a similarity here with the emergence doc-

trine of the relationship between the levels. Macroscopically described phenomena express a reality no less genuine than, but totally different from, the reality of the same ensemble microscopically viewed.<sup>6</sup>

Though the philosophical and gestalt approaches were broadly conceived, their focus proved too indefinite for our problem. As for the cultural and biological holistic theories, it was hard to know just what a "superorganic" entity *is*, or what it is that "emerges" as the reality at a higher level. Since it seemed to be conceded that consciousness, purposes, and values exist only in individuals, we are at a loss to understand what kind of transphenomenological existence or "macroscopic whole-character" the group whole may have to the cognitive experiences of its members (the parts). Even though we can agree that the whole is more than the sum of the parts, no one has been able to state in terms that are both denotational and general just what this "more than" or "different from" may be. We cannot expect this situation, however, to trouble a phenomenological theorist too greatly, since for him the property of denotability is not a uniquely significant criterion of reality.

EARLY GROUP INFLUENCE EXPERIMENTS—  
FRAME OF REFERENCE, SOCIAL NORM, AND  
REFERENCE GROUP

By 1924 the writer had published the results of the experiment assigned to him by Münsterberg. It should be stated that in these investigations, in the "together" sessions, the subjects did not communicate. They sat around a table in groups of from three to five, working independently, each individual having his own set of materials for the task in hand. From this work the following gen-

<sup>6</sup> Asch (1952b, pp. 265-267) makes some pertinent comments on the fact that many group (that is, interpersonal) happenings that are significant in the lives of individuals are happenings between *other* persons and lie outside the immediate knowledge or possibility of voluntary control by the individuals concerned. He also comments, by a "negatively" inferred method, on the control over individuals in a social situation as evidenced by the interfunctioning of social attitudes (pp. 577-579). These passages seem related to the present writer's theory of collective relations which will be developed later.



eralizations were forthcoming. Though these findings are here expressed as interpretations from data rather than as the data themselves, it should be borne in mind that each of them was based upon actual quantitative experiments.

Working in the presence of others, even though there is no direct contact nor communication, establishes certain fundamental attitudes. We are confused and distracted whenever we feel our reaction to be at variance with or inferior to the average behavior of those about us. In the association process we tend to inhibit egocentric trends and personal complexes. In our thinking we assume a conversational attitude, becoming more expansive and less precise. And finally, we avoid extremes in passing judgment, tending, often unconsciously, toward conformity with what we think to be the opinion of those about us.<sup>7</sup>

It should be stated that there was already some experimental background in the literature on group influence at the time of the writer's experiments; and in part his findings were confirmations of those of earlier investigators.

One can scarcely examine this record without a conviction that some kind of theory or generalization about "collective influence" is called for. For the present purpose attention is directed to the final sentence. The reference here is to experiments in which the subjects (graduate students in psychology) judged the degree of pleasantness of odors in a series, and the heaviness of a series of weights, in both the "social" and "solitary" situation. The results showed that in both these tasks they made *fewer* extreme judgments, that is, judgments toward either end of the scale, in the "together" condition than they did when working alone. This phenomenon of *group convergence* was considered by the writer at that time to be the indirect result of the adoption by the individual, when in the group, of an attitude to conform. That is, it was an evidence of a set, whether conscious or otherwise, to *avoid* the *nonconformity* to the (unknown) judgments of others that would be more likely to result from judging the odor or the weight to lie at either extreme of the scale than it would to result from judging it as nearer to the middle.

Some years later Sherif (1935), working

under different assumptions, reported a similar finding concerning judgments (together and alone) of the extent of the apparent movement of a stationary point of light in a dark room (autokinetic effect). Here, however, the judgments of the subjects were spoken aloud. Again there was found in the collective situation a marked convergence of judgments toward the central tendency of the subjects' estimates. The convergence effect, moreover, was seen to persist when the subjects were subsequently tested alone. These findings have been interpreted by Sherif and Sherif (1956) as due to "an extension into the social field" of a psychological principle that operates in perception and elsewhere. This is the principle that what is experienced by the individual in a situation is influenced by "anchorages" in some "frame of reference" afforded by that situation. Here the frame of reference for anchoring the subjects' judgments (they were called in this case "perceptions") consisted of the array of judgments given by the other individuals. Sherif's contribution to the problem of group and individual can thus be given in approximately the following terms. *A "frame of reference" is afforded by the situation of a reacting or interacting group in which the individual participates as a member; and this frame of reference provides an "anchorage" for the individual's perceptions (or other activities). Such an anchorage can be assumed to be a kind of collective standard or "norm" that arises among the individuals as a result of their interaction.* Since one would probably wish to exclude the hypothesis of mental telepathy, it seems to be implied in Sherif's logic that the reactions of the others must be overt and must be *perceived* in order for the collective standard or norm to arise. If they were not perceived, as in the writer's investigation, the effect, as we have seen, might be experimentally the same, but we would probably not interpret it as due to anchorage to a norm.

Be that as it may, from these and similar observations Sherif evolved his impressive theory of "social norms," applying the frame of reference principle not only to perceptions but to beliefs, to attitudes, and even to overt behaviors. As a corollary, one that relates the theory usefully to the phenomenon of levels, it

<sup>7</sup> Allport, F. H., *Social psychology*, p. 285. Reprinted by permission of the publishers, Houghton Mifflin and Company.



was conceived that a social norm, established originally by interaction at the overt, societal, level, becomes through continual use "interiorized" as a personal value or ideal of the individual himself.

In view of the fact that this effect was actually demonstrable only in the verbalized estimates of the (autokinetic) motion and not as an actual conscious perception, and more particularly since the present writer had previously obtained the same sort of result without any of the subjects knowing the judgments of the others, the hypothesis of the establishment in the group of a social norm, and especially of the anchoring of the perceptions of individuals to that norm, would seem overelaborate. That there was here an effect produced specifically by coacting with others was clear. But for the purpose of theorizing about the social realities involved, one could wish for a somewhat more parsimonious interpretation.

As an *a fortiori* principle, however, the theory of social norms has been very useful. By its very divergence from the writer's procedure, and by calling the influence in question a "norm" or "standard" instead of merely a "social influence," it did focus attention upon whatever in the situation is overt and capable of being conformed to, and upon the possibility that the character of the particular interactive situation (including especially the type of group involved) might affect the result. There might, in other words, be different kinds of norms for different groups; and, since the individual could participate in a number of groups, different norms for conformity might exist with relation to different phases of his activity as related to those different groups.

Once it was recognized that the individual might belong to a number of norm imposing groups, that is, that he might share group norms with others in many different contexts, a further important distinction arose. One could distinguish *different degrees of importance* to the individual of the collectivities to whose norms he was being "anchored." In particular, one could distinguish between the groups of which the individual was *merely* a member, and those with which, whether he was actually a member or not, he strove to

identify himself (that is, in which he wished to be *considered* as a member). Status classifications aspired to by the individual, to take only one instance, were added to "natural interest" groups as examples of this sort of affiliation. In order to distinguish groups of this special identification type, they have been called the *reference groups* of the individual concerned. *Reference groups, then, are those which provide the greatest satisfactions to the individual. They are the main anchorage for his values and beliefs, and their norms are reflected to a large extent in his attitudes, his standards, and even in his general behavior.* The concept of reference group affiliation has afforded a rationale for much provocative research.<sup>8</sup>

Although the social norm and frame of reference theory has proved attractive and has had some success in predicting behavior in collective situations, there are important matters bearing upon explanation which it leaves unsettled. For example, there is no generalized principle, beyond obvious teleological assumptions, stating what the norm represents in the nature of a strategic value to the collectivity. We are also not told the reason why the individuals conform so faithfully to it. Are group established anchors for perception any more effective than any other kinds of perceptual frames of reference? If so, why? Does the norm point only to something inside the individual, or is it a standard which he holds "on behalf" of a group; or is it implied that there is a distinct group reality in which the norm inheres? If the latter, what is the nature of that reality? Or again, does an individual perceive according to the norm because he actually believes there is a group which is imposing the norm on him, or is he implicitly agreeing, with others, to adopt the fiction that a group norm exists, and doing so for the conscious or unconscious purpose of

<sup>8</sup> We owe the concept of reference groups, as well as the term, to Hyman (1942). For a useful distinction regarding the use of reference groups, see also Kelley (1952). Newcomb (1950) has extended the idea by proposing that there are "negative" reference groups. These are the groups with which the individual wishes *not* to be identified ("negative" involvement), and whose norms and attitudes he therefore rejects or opposes.

using such a notion to control his fellow group members (as well as himself)?

The theory is concerned primarily with the fact of membership in or identification with a group, and with the "obligation" that is imposed (as if by the group) upon the members to perceive or act in certain ways. It is presumed, of course, that the individuals receive certain satisfactions through their membership, participation, and norm obedience. If the individual deviates from the norms and persists in doing so, he may find the rewards becoming increasingly uncertain and may eventually lose his membership in the group. In these respects the theory gives us a clue to the members' motivation for conformity to the norm. Our complaint, however, is that it fails to follow up the implications of the clue. Ought we not to consider more carefully how the so-called norms arise? Why, for example, did the original or "charter" members of a particular group come to espouse just these norms in the first place (as the group was being formed or thereafter) with the result that these particular standards and not others became the ones to be continually maintained and imposed upon all newcomers to this group. We are not so naive as to believe in a "group being" that gives them, or a "group mind" consensus that legislates them. There evidently must be some motivations on the part of *individuals*, motives not yet sufficiently explained as to their nature or their integrated action, that lead to norm creation as well as norm obedience. Let us return to the original experiments and see if we can formulate some hypotheses.

#### WHAT DID THE EARLY GROUP INFLUENCE EXPERIMENTS SHOW? INDICATIONS OF A PRIMITIVE LAW OF STRUCTURING—RE-INTERPRETATION OF SOCIAL NORMS

One of the reasons why the earlier experiments seemed so challenging was that there was no good a priori reason to expect the results obtained. The groups represented psychological laboratory sessions, a part of the regularly scheduled academic work of the leader and the subjects concerned. As thus constituted, it is hard to believe that these situations imposed any obligation regarding ways of perceiving or judging. The norm of

behavior expected on the part of the subjects in psychological work is usually just "to react naturally." Since there seems no other way of explaining these results on a group basis we are led to infer that, if such a basis did exist, it must have been because another (psychological) group had somehow arisen, interlaid as it were within the experimental groupings, just by virtue of the subjects being together and acting in one another's presence. There were, of course, the further facts that the subjects knew one another and that they were all doing the same kind of task. Such a hypothetical, situationally conditioned, group we shall suppose is the one within which the norm arose.

Now here we have a choice of at least two possible interpretations. Considering Sherif's (1935) experiment, we might first, as he seems to have done, conceive that the group was in some way established as a kind of causally prior condition. And then as the members announced their judgments, the norm emerged as a kind of estimated average (arrived at perhaps unconsciously) by each member, on the basis of the series of judgments he had heard expressed. Thus as the experiment proceeded, this subjective norm began to function psychologically as an anchor for the subjects' remaining judgments. The group (considered as here effective merely as a consensus of judgments) is the independent variable; and the direction of influence or control goes from it to the several members. And we infer that the influence is perceptual or cognitive: one's perceptions or judgments are actually modified by the evidence of what the others are experiencing.

This interpretation seems at least consistent with an example of alleged group determinism. Its chief drawback is that it lacks motivational dynamics. Either there is a "group entity" of some sort that acts to modify the individuals' reactions (an explanation one would probably reject), or the individuals do it themselves. If we choose the latter alternative, we are not told why they do it. Sherif (1935), of course, has said that "the group" has provided through the judgments of its members a frame of reference, and also, through its central tendency, an anchor. But frames of reference and anchors are not automatic. They have



to be of the individual's choosing in part at least; and we are not *told* why the individuals chose the group frame rather than some other. They might have chosen entirely, for example, their own earlier run of judgments (adaptation level), or those of some other single individual, or perhaps none at all. Furthermore, a direct effect of a *social* influence upon a perception of magnitude (if it is considered that the subject's report *does* describe an actual percept and not merely what he *says* he perceived) would be, as in all such cases, difficult to substantiate. There could be group effects upon judgments or spoken reactions without assuming the imposition of a norm of *perceiving*.

The second alternative, which in the writer's opinion involves fewer assumptions, is to consider that the norm (that is, the average of the judgments), instead of being a "one-way agency" of the group acting upon the individual, is circularly involved in the process of the formation of the group itself. The "group" depends on this norm no less than the norm upon the "group." In this case the (psychological) group, even if it is informal, sporadic, and perhaps fortuitous, is regarded as something whose formation and "very existence," depends upon the fact that it has a significant meaning to the individuals. Figuratively stated, it arises as a kind of theater of operations for the satisfaction of the individuals' needs, an implementation of their behavioral contacts, and a medium of their self-expression. It affords an opportunity that is generally (even if unconsciously) welcomed. It is against this background that the meaning of obedience to a "group norm" should be understood. The norm arises as a standard to be adhered to in that situation for a particular reason. This reason is that the relationship between the conduct which the norm prescribes and the cooperative activities of individuals which make the "group" possible is close and facilitating. The norm behavior is so positively related to the group and to its rewarding operations that it comes to be individually and spontaneously regarded by all as the effective (and *mandatory*) means by which the (valued) group condition can, and should, be maintained. A norm is thus not something that is "imposed upon" (or

even afforded to) the individuals by the group; it is one of the conditions of the existence of the group itself.

Let us canvass this situation in a little more detail, using as cases simple instances from daily life and the small co-working groupings of these earlier experiments. Unless there are special circumstances that determine otherwise, human beings when they are thrown together for any length of time develop sets or attitudes toward one another that reflect the perceptual realization of one another's presence and coactivity. Such sets may be scarcely conscious. They may give rise to words, signs, gestures of recognition, or similar "acknowledgments" of the other person; or they may remain purely covert. The subtle signs of the awareness of such interrelations are the most elementary forms of "rewards of participation" in such informal, scarcely conscious, types of groupings. They come to be expected, and in a sense, valued. At least one could say that their complete absence (that is, complete indifference of others to us) would probably be disturbing.

The satisfactions of an A in this situation come (back) to him in his perceptual recognition of the signs ("provisients") given off by B (or by a number of Bs), as A's "cycle" of behavior in its course toward "closure" becomes "tangent" to B's behavior in the environment. But we immediately note that the situation is mutual and reciprocal. B's rewards, under the usual boundary conditions of small groups, also come to him by way of some recognition given him through a stimulus afforded by A (or by a number of A's). When this happens, A and B are "bound together." That is, so long as this situation lasts they are more likely to stay in one another's presence or range of stimulation than to separate. If only *one* of them were to need or seek this sign of recognition or bonhomie from the other, the situation would be "linear," merely a case of A seeking and obtaining contact from B, but not B from A. There would thus be in the situation itself no guarantee of the likelihood or the recurrence of such contacts. The arrangement would therefore be momentary and undependable. But when each of them needs or seeks the other, be it ever so slightly or uncon-

sciously, there arises a predictively operating condition that we can call a "*collective structure*." We are justified in calling it that since its parts will not act, dependably and repetitively through time, alone. Through their *interdependence* the "ensemble" maintains that enduringly integrated character that we associate with the term structure.

It should be realized that it is really not the individuals as whole organisms, but *these particular give-and-take behaviors* of seeking and recognition, that can be most clearly said to constitute the group (that is, the "grouping"). And what is especially important for our purpose is the fact that this structure (A to B and B to A) *is self-closing or cyclical in character*. The give-and-take of two reciprocating individuals, in the manner above described, *is* a cycle in which the behavior of each receives closure from, and is bound through reciprocation to, the behavior of the other within a collective structure made up of the two.

In such episodes there is sometimes a *special* facilitating effect that arises because the individuals are not fully aware of the whole process. Probably neither person realizes in any deep or personal way the full extent of the other's involvement (desire for the rewards of the structure); hence the "binding" effect of the situation may exist (and exist in full force) without either being conscious of it. As each person is set to gain or retain the rewards from the other and also to pay the price of obtaining them by extending his own cordiality to the other, yet as each does not realize that the other also is motivated in exactly this same way, a state of special readiness which might be called "pluralistic ingratiation" becomes established. Such a condition can scarcely fail to be conducive to group formation or "group cohesiveness." Indeed we could almost say that the patterning of behaviors incident to, or resulting from, this binding effect *is* the group in question.<sup>9</sup>

<sup>9</sup> When we view human behavior, at either the individual or the collective level, in terms of its (self-closing) structure, important facts are seen which are not evident in the "linear-agent" view characteristic of molar stimulus response theory. In fact the whole problem of the ongoing and events of motivation, perception, and learning tends to take

Such, we may assume, was the state of affairs implicit in the co-working groups employed by Sherif (1935) and the present writer. One could expect that there was here in progress this subtle, perhaps unconscious, building up of self-closing structures among the individuals. Our earlier discussion of the results in this direction was limited to the quantitative convergence or conformity shown in perceptions or judgments made in the *together* situation. These results have been confirmed by other investigators. But in addition, other types of effects of co-working have been noted. Let us examine these to see if they also give evidence of the "structuration" effect before going on with our explanation of how norms arise.

The writer's researches (1920, 1924a, 1924b) included the performance of subjects working together, but individually, in small groups (as compared with the results of solitary work) in a number of activities. In one set of experiments the subjects did problems in multiplication, in another they looked at a reversible perspective figure and noted the number of fluctuations. Still other sets included the writing of words in free chain association, the responses being later classified by the experimenter as personal (that is, egocentric) or otherwise. The effect upon verbal reasoning was studied by having the subjects under the two conditions write brief essays on topics suggested by philosophical quotations.

Though some individual differences were shown in the results, the effects in general were as follows. In all situations in which speed and quantity of work (that is, amount of energy put forth) was taken as a variable more work was done in the group situation than while working alone. More problems were multiplied in the group, more reversals of perspective in the figure were noted, and more words were written in the association lists. This phenomenon was called by the writer "social facilitation." It was in part a confirmation of the work of earlier investigators. *Quality* of work, which could not always be graded objectively, did not seem to improve—perhaps the reverse. Errors made in multiplication on a different "format" and may be seen to exhibit a different set of laws.



plication problems, possibly through lapse of attention, were evidently more disturbing to the subject (as suggested by their bunching) than when alone; or else the distractions leading to the errors were of longer duration when others were present. In any case the presence of others seemed, to a considerable extent, and in one way or another, either challenging or preoccupying. Two other effects of much interest were noted. In the essays there seemed to be a more expansive conversational quality, with less profundity of thought, when they were written in the presence of others who were also writing on the same themes. (This effect, however, should be checked by other investigators because of small numbers of cases and the possible bias of the experimenter's judgment.) And there was in the group condition of the association experiments a definite tendency to associate (or at least to write down) fewer words that were in the context of a "personal" or subjective train of thinking.

Though sets toward others in the vicinity are subtle and difficult to formulate, it is believed that the above summary expresses the same general pattern of behavior that was seen in the norm and "conformity of judgment" experiments. Again there is suggested an effort to "belong with" or to "come to terms with" those around us. And again, the fact which is particularly interesting is that the situation did not really call for any such effort. There were no "terms with others" that were required by any overt circumstances. However *psychologically* imperative such tendencies may have been, there was nothing whatever to suggest to the subjects that they were *logically* called for. In these experiments the subjects did not even communicate. No comparisons whatever of the subjects' activities or productions were to be made at that time or later; and the subjects were well aware of that fact.

In spite of all these precautions an apparent drive to be "at one" with the others, or in some way acceptable or equal to the others, is an inference from the findings that is hard to escape. The increase in pace of mental activities on the part of most of the subjects in the co-working situation suggested a concern (perhaps one could almost call it an

anxiety or a fear) that they might "fall behind" or "fail to measure up to (their own or others') expectations." Some of the introspective reports were to the effect that "others are writing and I must write too." This increase of energy could be interpreted in various ways; but it does seem, in any case, to betoken an effort to adjust one's performance or "articulate" one's actions to the presence or actions of others, or to what one imagines the quantitative level of their activity to be. Even though one might prefer to attribute the socially facilitating effect to unconscious rivalry, level of aspiration, or ego feeling, the inference is still plausible that there is an attempt to adjust in one's own way to others; and that this effort actually requires the existence of some sort of group pattern. A dependable collective situation or ensemble is *necessary* for the expression of such trends of personality; otherwise these characteristics would have neither the milieu nor the standard of judgment needed for their exhibition and appraisal. In the conversational expansiveness of style and the reduction of personal associations in one's thought stream, the same tendency toward "interstructuring" with others appeared. Here again it seemed as though the subject were reacting for the benefit of present and watchful eyes, even though he knew that, though they were present, they were not watching; and in such a fashion as to create the impression (even though those so impressed would never be giving any evidence of it) that he is "outgoing," "convivial," and "fitting-in."

Possibly one could sum all this up as an intuitive effort for "security" among one's fellows, a striving so habitual and deep that it is represented in one's unconscious inner meanings even when the overt situation does not realistically call for it. From a more objective standpoint, the writer would prefer to call it simply a deep-seated tendency so to act as to establish some sort of a give-and-take, *structural*, relation with one's fellows—to be "not left adrift" but included as a part of a present collectivity; or to help create, if necessary, such a structural relationship so that one *can* be a part of it. Let the motives, interpreted consciously and teleologically, be what they will, let them be manifest or latent

—they all “feed into,” or “eventuate in,” the fact or existence of some sort of *collective structure* that is being formed, and formed with all the greater inevitability, perhaps, because its genesis is largely unconscious. Nothing could be gained by adding a new (structuring) “instinct” to the long and largely discredited list of instincts of the past. But where there seems to be such an inevitable result as this, may we not assume, as in all learned adjustments, that wherever the possibility of a ubiquitous and general means of fulfillment is present, probability alone (as in trial and error) might lead to its being employed. Collective structuring, in other words, could be considered as the result of the heightened probability of satisfactions through integrated or articulated behaviors, a probability that is afforded by the presence and potential interactions of others.

We can now return to the task of explaining how norms arise, and to the search for the basis of our explanation in these simple co-working or coexperiencing assemblies. In terms of the experiments, why did the mean of the estimates of the individuals’ perceptions in Sherif’s experiments come to be an anchor, so that the judgments converged toward it, giving a narrower spread of reactions in the group than when the subjects worked alone? And why was a similar convergence of judgments in the group situation found in the writer’s experiments on judgments of odor pleasantness and weight, even when the subjects did not communicate?

For the purpose of weighing the plausibility of the interpretations of so-called group effects that we have given in this section, the reader is now asked to put himself, imaginatively, in the place of the subjects in these experiments, and to test our assumption of “collective structuration” and its possible relation to these convergence phenomena. Suppose that the reader is judging with others the extent of apparent movement of the spot of light. He believes it to have been a very extensive or extreme movement; but the others, reporting before him, have all given estimates centering about a considerably smaller figure. If now he were to give his original “extreme” estimate, unmodified, would he not feel somewhat ill at ease? Might

he not, perhaps, even refrain from giving this estimate and substitute a smaller one? And if so, why?

Let us first take the perceptual social norm theory at face value. Here we explain the convergence tendency by saying that the social stimulation actually changed the subjects’ *conscious experience* of the light’s movement. There was a direct social effect upon the process of perceiving and the percept as such. We might interpret this effect, perhaps, as similar to what happens in cases of extreme suggestibility where the experience borders upon hallucination. But this is a no man’s land of psychology, too vague as to what is happening to afford good explanations. Besides, the really important question would still remain—why was the reader suggestible with respect to stimulation *from the group*? As an alternative explanation, the reader might say that though he “perceived” the movement as he did, in view of the strongly dissenting view of all the others, he might honestly think that he must be mistaken about its actual movement; and so he might tend to report a lesser movement to “correct his error.” This reaction would be uncooperative so far as the investigation was concerned; for it would be a clear violation of the experimenter’s instruction to report the movement of the light as it actually *looked*.

To take still another alternative, and usually a favored one, the reader might say that the reason why he would feel uncomfortable in giving the deviating judgment would be because it would make him appear conspicuous, different from the others, or “non-conforming,” or that it might either reveal the fact, or give some ground for the others to believe (truly or falsely), that he was a less competent, perhaps even a less intelligent, judge than the others. These reasons may be sound so far as they go, but are they complete? Why does one avoid being conspicuous in a group? Shall we attribute it to some sort of “instinct”? And what is behind this desire always to conform? Is this also an instinct, or does one do it because he feels that others expect it? But if this is so, why *should* they expect it? What does it do that is in some way baneful to their interrelationships or *modus vivendi* if they do *not* all act alike?



Besides one is not required to conform in everything he does in a group at all times. What then "selects" the behaviors on which conformity will be expected, and how is the pressure to conform imposed?

The explanation that the subject by changing his estimate is avoiding the implication of being less competent than the others, has, like the other explanations, an element of plausibility. But still, if he changes his estimate to save face, he really does not deceive *himself*. He can scarcely regard himself as a more competent perceiver than if he stuck to his original estimate, though he might conceivably convince himself that he really did not see it in the way he thought he had originally. But suppose he did suffer, under the "social influence," this sort of self-deception, or self-uncertainty. Would this not show a fairly strong tendency of susceptibility to what others are experiencing and the judgments they are making? And should this factor therefore not be fully as basic in our explanation of his behavior as his desire for vindication or "consonance" in his self-concept? And if the subject changed his estimate merely because he just wanted to *appear* competent, then we must at once recognize, as a required *prima-facie* addition to the "ego" motive, the important role of the *others*, that is, of the collective pattern of which the subject's activity is a part and in which his "appearing" is done.

In every case, therefore, we come back to the realization that the answers given above do not tell the whole story. Until we have found additional answers or some broader basis or hypothesis we must conclude that the convergence phenomenon is not yet fully explained.

Is there any alternative we have not tried? Yes, there is one. It lies in the direction we have earlier staked out and have hinted at in the analysis of ego motives just given, namely, that of the coming into existence of a psychologically grounded "collective structuring" of the individuals. Let the reader now begin again in a "negative" way by asking himself what the full consequences would be if he did *not* "converge" in his judgments toward the mean of the group. If instead of giving a similar response, he gave a *widely deviating*

reaction, what would be the effect upon others and upon their relationships with him. If one grants the premise that there is developing in the situation a relationship of give-and-take and a sense of "camaraderie" or "belonging," subtle and scarcely conscious, but nonetheless real, one might find in this very fact a natural explanation for the avoidance of the deviant response. Would the reason why the reader might feel ill at ease in giving a deviant reaction not be because such a reaction would run counter to these subtle attitudes or feelings of bonhomie that were established or were beginning to develop? Would it not be like spurning those with whom ties of fellowship were beginning to develop? Is it not at least a mild form of rejection to let it be known, even as a part of some routine occupation, that "I do not agree with you"?

Such a sense of the others we might well suppose to have been present even in the group situations used by the writer, in which the subjects did not communicate. To make an extreme judgment of an odor or a weight is to *run greater chances* of spurning or disregarding the judgments of one's associates, in effect, than if one makes a more moderate estimate. Psychologically the subjects were probably sensitive to one another's presence fairly continually, as the summary of the experiments and the introspections would indicate. Our own "probability" estimate of the appraisals, attitudes, expectations, and values of our fellows is with us no less truly than overt indications of these states on their part, and probably more continuously. One could easily imagine what would happen to the developing "we" feeling, or sense of solidarity, in a group situation if everyone sought to *diverge* from his fellows as much as possible in his judgments. And with the loss of these psychological components (cognitions or meanings) of being "bound" or united, that is, of being in readiness for reciprocally giving and receiving the rewards of cordial affiliation in the situation, the potentiality for what we have called the collective structure would be no more. One may well suppose that it would be on its way to extinction even if the diverging judgments were unexpressed but the

individuals were cognizant of their probable divergence.

Is this not, then, the answer to the problem of the norm? Is the norm not merely some behavioral stipulation or practice which if adhered to will be conducive to the formation, or continued existence and operation, of a structure in which the individuals have some degree of "*involvement*," that is, which they value and wish to preserve, a practice any marked deviation from which would tend to preclude or disrupt such a structure?

This conception also elucidates the matter of conformity, for which we found it difficult to assign a basic motive. Conformity, from this view, would be an *effect* and not a *cause*. Because individuals are similarly constituted, and those acts or omissions which would tend to foster or to disrupt their (valued) collective structure will be the same for all, the practices which they *individually* follow will be uniform throughout the collectivity. Their behavior with respect to the so-called norm, in other words, will be similar; and the belief may arise that this similarity is pursued and fostered as an end in itself. The individuals are said "to conform." Such conformity, however, in the primary condition, is merely the clustering of a statistical distribution under the influence of a common variable, and not a psychological fact or motive.

Traditional thinking to the contrary notwithstanding, the writer would maintain that *there is in human beings no such thing as a general and basic tendency, or drive, to conform. Nor is any systematic control or pressure toward conformity exerted by leaders, or by individuals upon one another, except as such uniformity of behavior is the result of pressure toward some other objective.* Whenever there is a pluralistic situation in which in order for an individual (or class of individuals) to perform some act (or have some experience) that he "desires" to perform (or for which he is "set") it is necessary that *another* person (or persons) perform certain acts (either similar or different and complementary to his own), we have what can be called a fact of collective structure. The structure is either collectively actualized or "potential," according to whether the desires are being carried out through the enabling action

of the other person, or remain merely covert as sets or meanings in the individuals concerned. Though this situation, in which a large number of persons are seen to be doing the same thing, or doing one or the other of two complementary types of things, has been called "conformity," such a term is misleading in that it wrongly implies some basic compulsion on the part of the individuals to do what other individuals are doing. It is as inaccurate, for example, to describe the action of two motorists in avoiding a collision as basically due to "conformity in obeying the rule of the road" as it would be to speak of the performance of the sex act as a matter of "conformity."

An individual really does not "perform a sex act." He performs certain behaviors. The sex act really represents (or contains) a fact of *structure*. The notion of human performance, since that means always performance of *individuals*, does not properly encompass it. It represents a juncture (or "event-region") of a collective structure that is composed of certain behaviors of two individuals. It is a reciprocating (that is, self-closing) articulating of the specific acts of two individuals, a structuring at the individual order, into one inclusive *collective* structure. A "marriage" consists of a mutually tangent *set* of such inclusive structures (of many sorts as to the "meanings" involved) articulated through event-contacts into a more complex (or "global") structuring of the behaviors of the two individuals (*à deux* collective structure). To call a marriage, or a family of husband and wife, a (dyadic) group is a loose way of speaking which gives the erroneous impression that the elements of the aggregate are total individuals (whole biological organisms and personalities) rather than the structured behaviors specifically involved. Such usage therefore conceals the true structural nature of the aggregate.

And so with every instance of a plurum in which a situation cannot be adequately described without tracing the interlocking or self-closing "juncturing" of the behaviors of individuals. This is the nature of *collective reality*. This is *structure* at the collective order. While dealing with this subject the writer would also express his conviction that



the above is a *general principle of nature*, that it appears at *all* levels, organismic, molecular, atomic, etc., as well as at the societal order. Thing, particle, organism, individual, and group—these are all agent-like terms which, like the “corporate fiction” of jurisprudence, are used as conveniences in our thinking. Wholenesses, or totalities, must be sought not in “things” or “agents,” for matter, as commonly conceived, does not provide a workable paradigm of wholeness. Nor do they lie in the notion of open-ended, linear, and “molarly” conceived acts of such agents. They lie, instead, in (completed or self-closed) structures of ongoing and events. Actually, we live in and through *structuring* at all levels; and it behooves us to try to understand its general forms and laws.

Conformity may of course be “worked up” into a secondary motive or principle of action; but its roots would probably reside as before in the desire to maintain a collective structure. A group leader may encourage the idea of a “group entity” or “agency” that is believed to be “imposing” such a norm upon the individuals and thus *demanding* conformity or obedience. By our hypothesis the fictitious character of the “conformity-demanding” referent of such thinking is evident. Psychologically, the fact that there *is such a way of thinking* is, of course, all too real. Individuals are *aided* in controlling others (and themselves), whether for good or for ill, through the belief that the standards they practice are the “rules” or “norms” actually laid down by a “superindividual” institution or a “society.” The belief in such a “norm giver,” or “obedience-enforcer,” as an actual being, is usually disclaimed when the individual is pressed to make a logical analysis. But it is worthy of note that there is a certain reluctance to do so; and almost never do we find an attempt being made to think out or explore the full realities that must be substituted if one is to understand what actually happens in the situation we call obedience to social or juristic norms.

In explaining the origin of norms as the discovery and confirmation of practices that preserve and facilitate the structure we have automatically thrown some light on why certain practices or dimensions of behavior are

“chosen” as norms and not others, that is, on the question of why norms are specific to the group and the situation at hand. To return once more to the earlier group experiments, we have suggested that wide deviation of judgments in regard to what the subjects were experiencing probably would have led to the loss of a feeling of camaraderie and to the breakdown of the potentiality for the structure. If this is so then the opposite trend, a clinging fairly close to the central tendency of judgment, would help to forestall this undesired outcome. Hence judging fairly closely to the level of the group average, actual or probable, as the experiment progressed, became the norm. On this basis one might predict what behaviors would be likely, in any group situation, to *become* norms. The introspections in the writer’s experiments reported, in connection with the facilitation effect, a certain feeling of obligation to keep working (keep writing) because the others were doing so. Again, such behavior facilitates the “structural” sense. If one had laid down his pencil and sat back, it would probably have been disturbing, no less to the individual himself than to the others.<sup>10</sup>

The multiplication task showed a possible tendency to become confused at what might have been the awareness that one is making mistakes that probably the others were not making. The collective structuring here leads to negative rather than positive self-feeling. It seems to imply an incipient, and perhaps unconscious, self-imposed standard of “accuracy” as a condition for “belonging with the others.” Or again, if in the writing of the essays or the associative trains of words the subject had lapsed into subjective reflections or soliloquy, he would probably have felt ill

<sup>10</sup> This observation introduces a new element which of course was present in all these group situations and was omitted in the discussion only to simplify our argument. There was an obligation of the subjects also (and perhaps primarily) to the experimenter to carry out his instructions. The experimenter also (that is, his behaviors and attitudes) must be considered as a part of the total collective structure—or rather, as representing a *special* structure with each subject which existed at times independently of the structure with the subject’s co-workers (when the subjects worked alone), and at times in tangency with the latter structure (in the *together* sessions).

at ease, as though he were "out of context." It would be as if the others, if they *could* know what he was thinking, would regard him as reclusive or withdrawn, and hence as "not participating in the structure." We thus see that the subjects, individually, were tacitly enacting what would probably have been overtly and explicitly stated as norms had circumstances warranted.

Here we have the structural version of the interiorization-exteriorization phenomenon so often recognized in systems of social psychology. One rather sharp realization, however, that comes in this connection is that the direction is here altered. There is an exteriorization *to* the collectivity, in our theory, from the "expansion" or environmental "actualization" of the structures of "meaning" within the individual's (interiorized) structure, rather than an effect in the reverse direction as social determinists have assumed. Structuring in other words must proceed from "lower" orders into "higher" orders of compounding. All this might well interest the social scientist as having a bearing on the question of social origins. Mores, customs, and other social norms, are as a rule unwritten, yet they are followed. They are sometimes considered by sociologists to go back to "time immemorial," and to be "vestiges" persisting by social inertia or lag. If our present view is correct, however, this is far from true. Instead, they should be thought to arise spontaneously; and their origin is not ancient but, in everything but a few formal details, continually present and immediate. Causation, in the structural view, is not historical nor linear, but continuous, time independent, and reciprocally cyclical. One looks for it neither in society *nor* in the individual, as traditionally seen as separate levels or agencies, but in the compounded patterns of structuring which are the essential reality underlying both.

Though we have traced the arising of norms as from interiorization to exteriorization, it should be noted that we are not coming back to individuals in the old telic-agent, or entity, sense. The same difficulty about referents for a "plurality of parts" that we have encountered in trying to find a referent for the group might face us in the last analysis with

respect to individuals. The conception we have arrived at is neither that of group nor of individuals in the older sense, but of the articulations or structurings of individuals' behaviors. These involve neither the total biological organism nor the (total) "personality" of the individual, but are patterns of "specialized segments" of behavior—cyclical acts and cyclical act sequences of such an explicit sort that through their "events of juncture" with the "act cycles" of other individuals they constitute the reciprocating, self-closing, patterns that we call the "individual" and the "social" orders. In all this, it will be noted that although we have discarded the group concept, at least in the older unanalyzed sense, we have by no means ignored the fact that there is a *collective* reality. This is represented by the fact that individuals' behaviors are not merely structured in and for themselves, but that they are also, in many instances, *interstructured* with one another to compose the more inclusive structures that we call collectivities. The collectivities also may in some cases be interstructured among themselves, composing assemblies of a still "higher" social order.

So much for the group experimental bases from which the reference frame and social norm theory were developed, and for our structural reconstruing of the evidence from that source. It remains now only to extend our methods to the larger organized, or institutional, situation within which also, the social norm theory has been elaborated. Many instances could be cited in which "conformity to a standard" takes on a more intelligible meaning when we see norm performance as a kind of behavioral structure or substructure that is, as it were, tangent to a certain collective structure, and necessary for the continuation of that structural pattern among the individuals concerned. Just as there are certain nonspecialized, convivial, "closures" gained in the smaller groups, (and indeed the group is predicated on just that fact) so in the operation of *any* enduring collectivity, such as an organization or institution, there are certain more specialized "rewards," the activities producing which bind the individuals together. And just as in the case of the small coacting group there arose certain implementing or facilitating behaviors, deviation from which



would have threatened this "cohesive" state of affairs, so in every enduring collectivity there are certain standards (or norms), arising through experience, obedience to which is conducive to the self-closing sequence of events that constitute that structure's operation, and violation of which tends to be destructive of that same cycle of events. And so, just as a statistical effect of "convergence" toward the mean of estimates appeared in the co-working situations, so uniformities of behavior in a more elaborate, edict obeying, or functional role sense appear in all the larger institutional or societal aggregations, the modal act in each case being uniquely adapted to, and protective of, the structure concerned.<sup>11</sup> Though deviation from the norm folkway, custom, or rule of law is frowned on and punished, the reason for this (like the reason for the sense of being ill at ease in deviations in experimental groups) lies not in the desire for conformity as such, but in

<sup>11</sup> A documenting of this statement and this paragraph as a whole in behavior in various social institutions can be found in the experimental program of the writer (Allport, 1934b, 1939; Allport & Solomon, 1939) and his students that led to the formulation of the "J-curve hypothesis of conforming behavior." Due partly to certain inadequacies in its presentation there were misunderstandings of the earlier statement of this theory on the part of social psychologists, especially as to the meaning of the "telic continuum" on which the observed behaviors were distributed. The matter can now be clarified by conceiving the J-curve of clustering (not "conforming") behaviors as fitted into place at the appropriate "event-regions" of space and time that are the site of the junctures in the collective structure concerned. An example would be the J-distribution of "degree of stopping" that occurs in a traffic intersection at the space-time region of a *stop sign*, where the juncturing is such as to permit the action of each motorist, in both of the two "directional" classes, to complete its cycle, without disruption, from the events of primary "need" (reason for making the trip) back to "closure" (that is, reduction of that "need" or "reason") consequent to the trip's completion. The norms we call traffic laws are simply experience-based, codified stipulations which, when practiced, enable these cyclical structures to complete themselves, and when deviated from tend to destroy (negate) these structures. The presence of J-curve distributions has also recently been demonstrated on large populations by Woo (1948) and the writer, in the practice of seven American male customs (norms) as related to the specific collective structures which they facilitate and protect.

its effect of diminishing the structure's operation, plus the damage or ultimate annihilation that might result to the structure itself should such deviations go unchallenged.

It is not difficult to envisage such hazards in any of our familiar collective structures. Disobedience with respect to a norm of table manners, for example, not merely mars a social occasion and subjects the deviator to reproof or perhaps to ostracism; it would, if persistently or widely practiced, spell the demise of the particular class distinction known as the "elite," or "polite society." With this would come the end of the usual give-and-take relations among those who now make up its personnel, the end of protocols and practices that are predicated upon distinctions of this sort. Failure of a man to earn a living for his family, and marital infidelity, are discouraged not only for the immediate interests of the parties involved, but because these breaches of norms might, not only in this particular case but generally, disrupt the *structures* of "matrimony" and the "family." To violate the norm of "freedom of the press" by instituting censorship would challenge not only the personality trends for "independence" in Americans, but the large and elaborate structures that constitute the news disseminating business. For church officials to cease stressing the norm that labels birth control a sin would not only go against the *personal* convictions of many church adherents; it might also weaken the significance to parishioners of those important substructures of ecclesiastic-parishioner behavior known as the sacraments. Not to stand when the "Star Spangled Banner" is played would bring reproof not merely because it shows a lack of patriotism on the part of the deviant, but because such behavior, if sufficiently multiplied, would result in uncertainty and distrust with respect to our personal security, since it would be a threat to the national structure of which our life-sustaining activities are an inalienable part. This list of norms and their bases is stated, it is true, partly in terms of all-or-none types of prescribed acts as well as in terms of the threshold degrees or quantities of action required. Perhaps, however, for that very reason they give the logic of our structural explanation a wider basis.

A further principle that is both interesting and important is implied in the preceding examples. There are usually two classes of rewards (or "provisients") of structures, or perhaps two substructural processes in close relation, that are safeguarded by norm obedience. The first is what might be called the regular or *in-course* rewards. These are the continuing outcomes of individuals' participations. They can be immediate; for example, there are the ego satisfactions derived as one mingles as a member in polite society, the newsman's profits in news gathering and publishing enterprises, the benefits and satisfactions, in marriage, of living with a mate, and the feelings of spiritual uplift experienced in religious worship. Or they can be delayed as in the case of the wages paid to the worker. The second class of rewards in structural participation are of two types. They are the indications that the structure *will be likely to continue its operation* (that is, tokens are given by the individuals to one another symbolizing their intent and readiness to continue playing their parts); and the indications given to an individual that *he is accepted* in the structure and that he will probably be able in the future to *hold his place* in it. These rewards, are, of course, more in the nature of meanings rather than of material goods. But it is usually important that the over-all provisiements, occurring in, or resulting from, the structure's operations, from time to time include such tokens of confidence in the structure. We might call these tokens, or the acts connected with them, *structural assurances*. They include formal signs, special phrases, shibboleths, and the like, that belong to certain circles and must be used upon certain occasions. They include also emblems of authority and membership validations, certifications in businesses and professions, symbols of national or governmental agencies, and tokens of the permanence and sanctity of religious institutions or personnel. Marriages and families with a feeling of unity and security among the members are not unlikely to be those in which anniversaries, birthdays, and similar occasions are remembered by suitable words or tokens. All these provisiements are like the oil in the oil gauge that tells us the engine is working

properly and will not be expected to break down.

The writer believes it to be one of the major supports of his hypothesis that so much is made of these "structural assurance" provisiements in collective aggregates and organizations. If the negative of this were true, the hypothesis would certainly seem doubtful. For such assurances, having always a future reference, underlie the *time independence* (endurance) of these structurings, without which there would be no truly stable human relations. Though their "dynamic" is secondary, that is, it comes mainly and in the long run from the regular *in-course* rewards of the structure, they are necessary for the continuance of those rewards. It should be noted also that they are *specific to them*. To try, for example, to interchange the structural assurance acts or tokens of polite society with those peculiar to government, church, or marriage, or to try to interchange any of the latter with one another, would be a *prima facie* absurdity. This way of putting it, however, brings out the structural or potentially structural character of collective facts.

Some assurance value of course attaches to the regular in-course provisiements of the structure; but there are usually also special structural assurance provisiements that are required. In certain cases there are whole *substructures* of reciprocal and formal token giving and receiving acts that are "tangent," as it were, to the main in-course cycle of the total structure or to other "assurance cycles." There are also corresponding "assurance needs," (primary upsets or disequilibria in individuals) that are attendant upon this process and must be occasionally or periodically satisfied through the "closure" of the formal substructure concerned. Especially important for our present purpose is the fact that such assurance providing substructures often become the focus of norms that are adhered to rigidly, often fanatically. Having the visibility of "oil gauges" they become the stipulations for efforts to create, preserve, and extend the structural bases concerned. So-called obedience to a norm, whether it be a norm of in-course participation or of structural assurance, is as truly, and even more basically, a *cause* or *support* of group ex-

8-430

8.9.70



istence and group action than it is an effect. We come nearer to understanding it when we consider it as continuing structural phenomenon (or structurizing process) rather than a manifestation of a "social influence" on perception or behavior, or an aspect of group determinism in the traditional and unanalyzed sense of the term.

The normative, perceptual, and cognitive systems of social psychology have done much, both experimentally and theoretically, to bring these important matters to attention and to show that collectivities, however they may be described, have a real existence and should be studied. A logical analysis of certain fictions, however, and a more intimate observational approach to what actually happens between or among individuals, are needed to bring us closer to a solution of the master problem. When these aids are enlisted one is able to reconstrue the facts so well adduced by the phenomenologically and normatively oriented theories into a more objective and explicit conceptual model. In this model not merely percepts, subjective anchors, and values, but the actual patterning of intra-individual events, producing collectivities in a more overt and physicalistic sense can be brought to bear.

We turn now to other theories. In the one next to be discussed the answer to the master problem is given through a revived version of the more traditional notion of the group, in which a collectivity is conceived not merely as a norm giver, but as a truly dynamic and behavior determining agency.

#### FIELD THEORY AND GROUP DYNAMICS— STRUCTURAL REINTERPRETATIONS

Although frame of reference theory leaned somewhat toward social determinism as the mean of individuals' perceptions, it did not elaborate the group concept by ascribing to it special dynamic processes or laws. It was but a step to this postulate however, and its development was spurred by the practical usefulness of such a view. If the investigator fixes his attention upon collective episodes wholly within a group frame of reference (without being too hesitant about using collective fictions), he may be able to formulate a science of group action or group perform-

ance, that is, a system having quantitative functional relations between variables stated entirely at the collective level. Such a program has actually been carried out in the movement known as *group dynamics*.

Any experimental program in social psychology will be likely to be based upon some background theory about individuals and groups. Serving this purpose for group dynamics, at least at the start, was the psychological field theory originated by Kurt Lewin (1936, 1947). The group reality was conveniently represented in this theory in the form of a topological field (life-space) with its boundaries, subregions, goals with their positive or negative valences, and vectors of "psychological forces" acting upon individuals. The life-spaces of the individuals overlapped and, taken together, represented the life-space of the group. The answer which this theory has given to the "master problem" is stated in phenomenological rather than physicalistic terms, as follows: *The individual is a point within the life-space (field) of the group. This field is under tension due to (unbalanced) social forces acting in it; and the individual is pulled this way and that by the resultant of whatever forces are operative, as the field seeks to come to equilibrium.* The forces were conceived as related to tensions within the individual. There was also said to be a "locomotion" of the individual (or group) through intervening regions toward a goal (or away from it if the goal was negative). The distinction between the individual and group in respect to life space, motivation, and locomotion was often blurred. It was not clear whether the individual goals were also the group goal, that is, as the combined or similar objectives of so many individuals, or whether the "group as a whole" was conceived as having a separate goal of its own, resulting from the integration of the individuals' different goal directed behaviors. In the latter case, the problem of understanding, in terms of a single field model, the interrelationships of the individuals' goal directed activities, or of representing in a single diagram both the individual and group levels of action, was even more baffling. In spite of these handicaps the model has had great vogue and has

led to many useful interpretations and many experiments.<sup>12</sup>

Started by Lewin (1947, 1951) and systematized and documented by Cartwright and Zander (1953), group dynamics has been a discipline mainly of relatively small groups. This limitation was necessary from both an experimental and a theoretical standpoint, since the group situations had to be capable of being controlled. Hence the relation of individuals to one another in "institutions," or in the "great society," had to be correspondingly slighted.

The topological field theory was operationalized for experimental purposes by a further analogy. One can "interpret" the theory of forces in a group field by measuring the social outcomes that are said to be due to these forces. Since the outcomes are assumed to be proportional to the strength of the forces, such measures will serve to quantify the forces involved. *The social forces within the group represent pressures that are due to "field conditions" of group organization or process, inherent or imposed. The pooled statistics of the behaviors of the individual members (acting under such field conditions) are taken as measures of the "locomotion of the group" that is produced by these pressures.* For such a program, the name group dynamics seems fairly appropriate. The field conditions exerting the "pressures," and thus representing the *independent* variables, are exemplified by such group influences as "solidarity" or "cohesion," "group goals," place of residence, competition or cooperation, type of leadership, "social atmospheres," "social climates," and the "importance of an issue" to the group. Among the *dependent* variables ("group locomotions") one might mention attitudes, productions (quantity and quality), amount of communication, discussion directed toward changing the opinions of dissident members, conflicts, acceptance or rejection of individuals as members, and the adoption of common standards. A number of such hypotheses has been formulated and tested by Festinger and his colleagues (1950). Within this general framework much has been learned

about the effects of these independent "group variables" upon these "group outcomes" or "locomotions." Both the program and the theoretical model of group dynamics, however, are broader than the above formulation would imply. Cartwright and Zander (1953) have republished and summarized examples of various ideologies and methods. For the general subject of "behavior in groups" one can consult Hare, Borgatta, and Bales (1955) and Thibaut and Kelley (1960).

In group dynamics as in other theories of social psychology we become acutely aware of the problem of the levels (group and individual) and of the dilemmas that come from being compelled to take some sort of stand regarding them. As pointed out earlier, if one talks about the group but does so in terms that apply also to individuals (as one naturally must in dealing with the dependent *experimental* variables), one must expect to run into tautologies. Here the using of the concept of a group in connection with the description adds no new content or meaning. The same is often true when one speaks of group conditions, as when an investigator speaks of groups under frustration or tension. In other instances the invoking of individual terminology may lead to group metaphors which, when actually applied to the acts or interests of all the individuals, are not truly designative. On the other hand, there are cases occurring in group dynamics, as elsewhere, in which, if one *does* concede that the description (that is, the variable to be measured) *is* something that a *group might properly be said to do*, or a property that a group *could have*, there is then at hand no denotable evidence of any group that is doing this act or having this property. This is the case when one speaks of social norms, of group pressures, or of cohesiveness, as a property of a group. What is here *predicated* of the "agent" seems appropriate; but the agent itself, as an actual doer of the act, or a denotable possessor of the property, is missing. We see no group that is "pressing" on the individuals, or is "sticking together."

When the group dynamicist speaks of the "attraction of the group for the individual" does he not mean just the attraction of the individuals for one another? If individuals are

<sup>12</sup> For an instructive diagrammatic employment of the model as related to group dynamics, see French (1941).



all drawn toward one another, are they not *ipso facto* drawn to the group? Or, on the other hand, does the group dynamicist mean the attractiveness to the individuals of some complex of interrelationships, or of the interactions of each of them with each of the others, by virtue of which they are all classified as members of the group? Does he define "attraction" to mean, *inter alia*, the possession of some common interest or common bond? If such are his meanings, and if this complex of interactions, or the common interest, is not kept at the focus of attention and perhaps specified when the term attraction is used, then an ambiguity is likely to be fostered. If it is to be specified, we shall then probably have to recite a whole nexus of meanings (embracing various individuals) so complex as to make the notion of attraction to the group too inexplicit to serve as an experimental variable.

But we need at this point to clarify our position. The exponent of the theory may claim that for his purpose it makes no difference just what form our definitions take so long as we are consistent. All that he is trying to do, he will say, is to take some set of operations which *can be called* "measuring the attractiveness of the group" and relate them experimentally to other measurable variables. If he can do this, then he can say that he has achieved a useful purpose in making predictions possible; and he has, moreover, been able to do so without going into the minutiae of exactly what the term "attractiveness" means. One could cite as a parallel case the situation in S-R learning theory in which terms are invented whose full significance is not understood (reinforcement, drive, habit strength, etc.), but which facilitate operations of measuring and relating variables, operations which themselves are never vague or in doubt as to their meanings.

With this the writer has no serious quarrel. He is only saying that there are other objectives also for which concepts need to be defined. Though it is not essential to the prediction and control of outcomes of collective action that we subject such terms as group pressures, group goals, group attraction, and group cohesiveness to semantic scrutiny, it is necessary that we do so if we are to solve

the problem to which we are here committed. The exponents of group dynamics have devised their operations and their experiments at the individual level. This has been necessary because groups have for us no denotation independent of that of individuals. But they have then proceeded to talk about these matters at the group level. The question we are asking is this: What is the value of such a way of talking when we consider it not for the purpose of predicting group variables or outcomes significant for group management, but for the purpose of understanding what is happening as these changes in variables appear?

Let us examine the alleged parallel from learning theory. In the first place it must be noted that the cases are not strictly comparable. In learning theory, as pointed out in the opening section of this article, there was a much better denotational base (the rat) for postulating some nondenotable variable (such as habit strength) than existed in the group formulations for the group related variables; and this basis made it possible to attain better control and precision in experimenting. But even if this were not true, the citing of what is done in research in learning does still not justify the neglect of the more careful analysis of concepts that is required for basic formulations. In the case of learning theory though it is true that the operations and data which can be rationalized by such terms as "reinforcement" or "habit strength" lead, in part because they are so rationalized, to useful laws concerning the quantity of learning achieved under different conditions, can anyone seriously say that these concepts, both of them metaphorical, or even the quantitative laws into which they help to order the data, are adequate descriptions of the occurrences taking place in an organism as the organism "learns"? If learning is to be completely understood, these molar concepts of learning in which the process is seen "from the outside" must ultimately be related in some way to neuronc assemblies, synaptic and humoral transmission, and the relation of "information" to molecular processes. Even the most thoroughgoing molar psychologist would probably not be such an "isolationist" as to deny this statement. Is it not so, also, when we try

to extrapolate our "individual" observations to description at the "group" level? If the antireduction experimentalist cannot deny at least the *relevance* of structurings at the organic level (that is, within the individual) to the level of description in which he is interested, can the group dynamicist, who must use the lower level elements (individuals) *to work with* or he will have nothing, escape in his account of group processes (group attractiveness, pressures, locomotion, and the like) the responsibility for describing what these individuals are doing as all this group activity takes place? Until the individual and the social levels have been tied together through some clear and systematic theory, a task that has not been accomplished in group dynamics, we must expect our social science to be *ad hoc* rather than distinterested and rigorous.

Even from their own pragmatic position there seems to be some need of caution lest predictions go awry. We may ask, first, what group sampling methods have been used for ascertaining the "laws of groups." Though there have been various degrees of permanence and stability in the group situations employed, such data and indices are usually obtained from the population of a small aggregate, under fairly specific conditions, sometimes involving a definite place basis or a set of particular circumstances.

Then too, the problem of "partial inclusion" becomes important. A familiar method in group dynamics experiments is to assemble and treat the data in group-wise fashion. Means of individuals, percentages, numbers of cases in certain categories, and the like, have been computed as indices of measurement of the group variable. The fact is not experimentally provided for that the individuals from whose behavior the "group laws" have to be derived belong also to other groups. The same individual can belong to many. What the individual does in one group, or merely his relation to that group, may have an important bearing upon what he does in another group; and the total "group membership manifold" of one individual who is a member of a particular group may be widely different from the manifolds of the other members. Personality characteristics of individuals

may also be related both to their choices of groups and to some degree to their behavior in those groups. How tenuous, shifting, and contingent group entities turn out to be!

If one wishes to say that although these things are true there is still in all groups a *residue of sameness* (other variables being held constant) just because they *are* groups (and since we are seeking laws of "the group" *as such*, we are interested *only* in these same-nesses) or if one says that he wishes to consider a group, by definition, as a thing only of the moment, a "gathering"—then it is true that we do not need to worry about the other group involvements or the personality characteristics of the individuals who compose it. But in so limiting our problem we would have to be reconciled to overlooking not only group differences, uniquenesses, and inconsistencies, but a great deal of the significance of "group affiliation" in the lives of human beings. A "pure" collective level, with laws all to itself, might be a pretty arid thing. Laws of group dynamics that work perfectly at such levels might turn out to be more or less limited by the investigators' selection and methodology.

What we are dealing with as the group reality, both in simple co-working situations and in aggregations with high affiliative involvement, is, as has been previously noted, more like "segments" of individuals' behavior patterns that have become in one degree or another organized with similar segments of the behaviors of others, than it is like an aggregate of whole individuals. The individual himself is a sort of matrix in which these patterns of collectively organized segments meet and affect one another. Instead of saying that a group incorporates (or is composed of) many individuals, we would do almost better to say that an individual incorporates many groups. One group has salience for him (that is, he is present in it or acting in terms of it) at one time, and another has salience at another time. When not salient for the individual, a group to which he belongs could be said to be represented in his own organism as sets, latent meanings, or stored memories.

Though the problem of partial inclusion has been recognized by workers in group dynamics (Cartwright & Zander, 1953, p. 146; Fest-



inger, 1950), they do not seem to have found a solution for it in terms of their own system. Then too, other types of questions that have arisen, and have been noted by these investigators themselves, betray a similar lack of clarity in the line which separates the ontology of individuals from that of groups. Who, for example, exerts the pressure upon individual members for conformity? Since this is perhaps seldom the work of a sole specific agent or group monitor, we must seek an additional explanation. Do the individual members coerce *one another* (and themselves)? If so, by what process do they do it? Or again, how can one explain the tenacity of group standards? Under what circumstances do they change? Or, when we say that certain questions for discussion are important for the *group*, what do we mean? Does this mean anything more than "important for the individuals"? If so, how do we spell it out? By what process does an individual goal become a group goal? What sense does it make to speak of a group goal becoming an individual goal? Was not the group goal always a goal of individuals? In such questions we see the logical difficulties attending almost any form of societal determinism.

It is suggested that these dilemmas arise from trying to treat an hypostatized entity (group) that cannot be uniquely denoted as though it *could* be denoted. Can we not resolve the difficulty by a different approach? The writer would again offer for this purpose the interpretation of a collectivity as a reciprocating structuring of the segments of individuals' behaviors. But we shall here first elaborate that conception a little more in quasigeometric terms. The collective or group level of structure is made up of, and only of, these acts of individuals. Such behaviors of individuals separately are in turn structures at a lower (individual) order. The model is thus self-closing, or cyclical, throughout and at all levels. It starts with an "upset," or events in a "primary event region" (disequilibrium, "need," etc.), in an individual, and returns through the sequence of actions of all concerned, including the individual himself, back to the original "region of upset" in that individual. This return of acts to the primary event region, which is called

"closure," changes the earlier state of affairs back toward equilibrium. It should be noted that if the collectivity is constant and enduring (as groups more or less are), each individual in it will probably be seen to be a possible "closure person." The model holds not only for two individuals but for any number. To mention the full geometric possibilities in the latter case would take us too far afield. But the structure, in any case, can readily be shown to be cyclical and self-closing and, as such, can be illustrated in aggregates of any size or degree of elaboration.

It will be remembered that the building of the informal or the co-working group (structure) was pictured as the beginnings of the operation of a self-closing sequence of acts and rewards of "bonhomie," involving an individual's segments of behavior "reciprocating" with those of others. Such self-closing sequences were said to be subtly valued or desired by the individuals. We pointed out that that which often *looks like* "group determinism," "direction," or "group pressure," may not really be such, but may be better understood as the things which the individuals are doing to establish or preserve just this cyclism of interactions and their rewards. What looks like "group controls" are really structurizing behaviors of individuals. We are dealing with a process of *collective structurization* of individuals' behaviors that *eventuates* in establishing or preserving the "group" (that is, the structure), rather than with a process of the determination of individual behavior *by* a group. The term group "influence," or the idea of the influence of the group *upon* the individual, is therefore a misnomer or a misconception. The laws we are concerned with, by this hypothesis, are laws of "structurogenesis" in the collective human sphere, *not* laws of "social agency" or the "dynamics of groups."

The writer concedes that this interpretation seems to apply more directly to such phases as structure *formation* and *assurance* rewards than to regular in-course operations of an "established" structure. The principle, however, will still hold in a broad sense, and this interpretation will help us to translate the working program of group dynamics into a somewhat more rational form.

Let us now take the questions quoted earlier as presenting dilemmas for this program. The question of who exerts the pressure to conform has already been answered, both here and in the discussion of norms, by pointing out that there *is* no such pressure. What looks like it is really only a narrowed statistical distribution that results from the fact that acting in a certain way in a certain situation is necessary on the part of all the (similarly constituted) individuals in order for a structure that was establishing itself among their segments of behaviors to be formed and survive. And when we speak of a discussion question that is "important for the group," we could, in this view, mean a question that closely bears upon this structuration process. We could mean the bearing of the acts of individuals, as specific events, on the collective structure *per se*, including both its operational (or in-course) provisients and its structural assurances. Considerations which we might call the "geometry of the structure" could here be important. In order to translate into structural concepts the notion of a group goal we would abandon the linear teleological implication of goal and substitute the cyclically accruing "provisient output" (material or nonmaterial) of the structure concerned. Where the individual goal is said to be identical with the group goal this phenomenon can be explained by the hypothesis that there is structure of meaning (meaning cycle) in the individual's organism that is, more or less, a conceptual replication of the *collective* structure concerned, a structure of which his own behaviors are a part. This is the event-structural interpretation of "exteriorization and interiorization." We must remember also that "partial inclusion," which is explained by the fact that the structure is one of "segmentalized" reactions and not anatomical persons, permits the individual many such group goal replications (identifications), and also permits that some of his organismic structures (as for example personality trends) are not in *any* collective structurization. The individual, we have said, is in part a "matrix," a relating (or tangency) in his own structural system of the many collective structures in which his (segments of) behavior play a part. Partial inclusion, it will be noted, is capable

of fairly clear formulation in terms of the structural hypothesis.

Finally, for group standards or prescriptions, a generalizing of our interpretation of norms will provide the needed clarification. We can think more broadly of the norm as typifying any structure of behavior that is tangent to a main collective structure. As we have seen, such tangent structures often have the probability of being either facilitating or inhibiting to the collective structure under consideration. The behaviors involved in such tangent structures as are facilitating represent the "positive" prescriptions, that is, the things which, one, as a participant in the structure, will be motivated to do, or the "quantities" or "degrees" in which one must act; while the behaviors inhibiting tangent structures become the "negative" prescriptions, the things one must not do. This further development of the topic opens up the employment of a basic concept that we can call *relevance* (really "interstructurance," in more rigorous terms); and through this notion we can relate not merely norm (structures) to the main collective structure, but any structure whatsoever to any other structure upon which it may impinge. Such interrelationships are of course, fundamental to motivation for (or in) a particular structure. The "intensity" of our behavior in a given structure is in part dependent upon these reinforcing or inhibiting effects received from positively or negatively related structures. The interstructurance relationship can be readily mathematized as a *relevance* (or *interstructurance*) *ratio*. It can have either a +, a -, or a 0 value. And as the reader will recognize, it is also a matter of degree. Certain tangent structures are *more* relevant (positively or negatively) to a main structure than certain other tangent structures; and the size of these relative ratios can be estimated upon a scale or otherwise empirically measured. Attempts to formulate and use such indices experimentally have met with success in the investigations of the writer and his students.

Group standards, therefore, can be conceived as tangent structures of stipulative, communicative, acts (or potential tangent structures represented in individuals' meanings) which have high positive relevance to



(interstructurance with) the main operational structuring of the collectivity concerned. They can also be conceived as "substructures" of this main structure. Circumstances, deliberate or fortuitous, may change the degrees, or perhaps even the sign, of their relevance to that structure. When that occurs the group standard will change.

As previously demonstrated, the structural cartography makes the behaviors of individuals integral (and integrated) parts of the group model. They must, in principle, be *shown*, and in their precise articulation with each other mapped out, or else no group (collective structure) can be either depicted or conceived. This consideration plays an important part in the translation into structural thinking of the experimental logic of group dynamics. Because of the conceptual vagueness and unsteadiness of the group as a working notion we have earlier expressed the wish for some other procedure than group indices as a method for the accumulation and treatment of the data. Such a basis is now afforded by the behaviors of *individuals*.

For an example let us turn to a question that has been central for group dynamics, namely, the definition of "group cohesiveness" as an independent variable, and the question as to how it can be measured. This task has been performed through a number of group indices, such as mean numbers of friendly comments or of friendships among the members, degree of the sharing of norms, and mean avoidance of absenteeism. For none of these methods do we seem to have any sure means of its validation as a measure of just this elusive variable.

To obtain a translation of the concept into structure theory terms let us go back to the early group experiments and seek its basis in connection with those findings. According to the interpretation we have given, the individuals, tacitly, perhaps unconsciously, desire the collective situation (structure) to come into existence or to endure. They receive provisients and closure from it; they have in it a definite interest and involvement. The one important quantitative variable we have not thus far exploited to any extent in our theory is the variable which is suggested by the question: "*how great* is that interest or involvement?" And of course we must mean

"how great is it for the individual" (that is, for each participant)? This variable, as an index, expresses the concept we have called the *potency of involvement* of the individual in the structure concerned, or, in more general terms, his *index of structurance*. Its value will no doubt depend upon the number of provisients and the general amount of "closure" the individual receives through the structure's operation. It might also represent the amount of energy that goes into the individual's participating acts. For estimating it the writer has had success with a method of "equivalents" which he has called the "negative causation" method, which will presently be described. It is this structural index, then, the "structural quantics" or energy, or the "potency of involvement," that the writer offers as a basis for determining the equivalent in structural terms of group cohesiveness.

It is conceded that such an index may be more difficult to operationalize with certainty than the group-wise method of counting overt friendly contacts, tabulating instances of absenteeism, and the like. The following procedure used in the experiments of the writer and his students is offered merely as a suggestion. The logic of the concept of potency of involvement and the possibility of its role as a structural variable does not depend upon the employment of this particular method. In this procedure one asks the subject (on suitable questionnaire forms) to imagine that the group concerned (that is, one of the structures in which he is involved) is threatened with dissolution, or that he is in danger of losing his place in it. He is then asked what proportion of his spare time (or energy) he would be willing to give under these circumstances to "keep it going" or to "maintain his membership" in it. In individual terms, "cohesiveness of the group" is thus translated into "potency of involvement of the individual" in the structure. From those who hear of this method for the first time the almost inevitable reaction is that it is too introspective or subjective to be reliable. We shall not undertake to defend this point. The question here is merely whether, in the absence of an opportunity for more overt indications, this less than perfect method is better than no attempt to measure potency of involvement at all. In defense of the procedure we would note that the subject

is not being asked to rate or appraise himself in any way; there is no apparent reason for any differential bias or evasive answering; and there is, in general, no one who could be expected to know better than the subject himself how important the group (that is, the structure) in question is to him. Still more to the point is the fact that, almost without exception, clear-cut positive results have been obtained upon hypotheses dealing with a considerable range of behaviors in which potency of involvement, measured in the way above indicated, was used as the basis of the independent variable.

If one grants that a measure of potency of involvement that is dependable can be made available, would not such a measure, when combined for all the members of a given group, represent in terms of individual meanings just what the group dynamicist is seeking as the referent of his group term "cohesiveness?" Such a "potency of involvement" of an individual in a group (structure) is probably not a simple thing. We would always regard our index (subject's answer to our question) as embodying also the effect of whatever increments of motivation (positive or negative) are accruing to the set to preserve the structure in question by reason of the subject's involvement in *other* (tangent) structures. For we remember that the individual is, in these respects, like a matrix of relationships of many groupings. We should also consider the individual's index of potency of involvement in a certain collective structure as embodying the effect of motivational increments, again positive *or* negative, that are received from his characteristics (inner "meaning" structures) of personality. There would thus be assumed to be attained in the organism a kind of algebraic summing of the effects from all these other sources, as well as from the structure immediately under investigation, to produce the individual's "net investment" in, or tendency to maintain, the collective structure concerned. Would not such a measure, when the individuals' indices are averaged (or totaled), give a clear operationalizing of the Festinger concept of group cohesiveness? What index could *better* represent "the resultant of all the forces acting upon all the individuals to leave, or to remain in," the group concerned?

With potency of involvement and relevance (that is, structurance and interstructurance) clearly defined and measured as variables, and applied to structural magnitudes and relations in all the different levels of structuring in an individual's (inner and outer) manifold, we now have before us the entire motivational basis of the individual as seen from the structural standpoint. We have in hand, in other words, the equipment for predicting the amount (or energies) in which almost any act structure typical of the individual's behavior may be expected to be performed. This means that we should be able to predict the amount of any dependent variable activity performed singly or jointly with others, such as the amount of discussion, "distortion" effects, acceptance or rejection of a proposal, attitudes, morale, conflicts, effort for consonance, energy put into any single or joint project, or any of the other dependent variables used in group dynamics or elsewhere. By weighting the amount of the subject's potency of involvement in all the structures of the subject's manifold having appreciable relevance to the behavior (structure) to be predicted, by the relevance ratios (+ or -) of those structures to the behavior (structure) to be predicted, and then summing all these increments algebraically, the independent variable for making such a prediction can be obtained. The sum or mean of such predictions for all the individuals in a group should provide one with the group index for the activity to be predicted, as dependent upon the combination of collective structures in the entire individual manifolds of the population making up the group from which the data were obtained.

This procedure would afford a prediction that *does* take into account all the other involvements of the individuals who happen to be in that group. The immediate group (structure) concerned, however, since it would be more "salient," might afford a greater increment of weight to the behavior to be predicted than other groups. Increments coming from relevance to *particular* groups (structures) can be separately studied as desired. If one wishes to find the strength of the *total* gamut of motivation, the increments from relevant personality trends (structures that are *in* the subject) are to be included.



One particular advantage of this translation of group dynamics objectives into structural methods is that the method is not limited to small groups but can be used upon populations in larger collective aggregations, or institutions, of any size or scope.

The details and symbols of the "structural quantics" equation explained verbally above have been published elsewhere by the writer (1954, 1955). To the reader is left the task of envisaging its application to the prediction of individuals' participations in significant action programs or attitude commitments, between which actions and commitments and the collectivities in which they are involved there is appreciable interstructure. A résumé of experimental work already done in this field will be published as soon as practicable. Judging from the positive results already attained, there seems reason to hope that this approach, through the aid of the event-structure principle, may be able to generalize and unite in parsimonious fashion not only hypotheses of group dynamics but the findings of many other classical experiments in social psychology.

#### REFERENCES

- ALLPORT, F. H. The influence of the group upon association and thought. *J. exp. Psychol.*, 1920, 3, 159-182.
- ALLPORT, F. H. The group fallacy in relation to social science. *J. abnorm. soc. Psychol.*, 1924, 19, 60-73. (a)
- ALLPORT, F. H. *Social psychology*. Boston: Houghton Mifflin, 1924. (b)
- ALLPORT, F. H. "Group" and "institution" as concepts in a natural science of social phenomena. *Publ. Amer. Sociol. Soc.*, 1927, 22, 83-100. (a)
- ALLPORT, F. H. The psychological nature of political structure. *Amer. pol. sci. Rev.*, 1927, 21, 611-618. (b)
- ALLPORT, F. H. *Institutional behavior*. Chapel Hill: Univer. North Carolina Press, 1933.
- ALLPORT, F. H. Individuals and their human environment. *Proc. Ass. Res. nerv. ment. Dis.*, 1934, 14, 234-252. (a)
- ALLPORT, F. H. The J-curve hypothesis of conforming behavior. *J. soc. Psychol.*, 1934, 5, 141-182. (b)
- ALLPORT, F. H. Toward a science of public opinion. *Publ. opin. Quart.*, 1937, 1, 7-23.
- ALLPORT, F. H. Rule and custom as individual variations of behavior distributed upon a continuum of conformity. *Amer. J. Sociol.*, 1939, 44, 897-921.
- ALLPORT, F. H. Polls and the science of public opinion. *Publ. opin. Quart.*, 1940, 4, 249-257.
- ALLPORT, F. H. Methods in the study of collective action phenomena. *J. soc. Psychol.*, 1942, 15, 165-185.
- ALLPORT, F. H. The structuring of events: Outline of a general theory with applications to psychology. *Psychol. Rev.*, 1954, 61, 281-303.
- ALLPORT, F. H. *Theories of perception and the concept of structure*. New York: Wiley, 1955.
- ALLPORT, F. H. The contemporary appraisal of an old problem. *Contemp. Psychol.*, 1961, 6, 195-196.
- ALLPORT, F. H., & HARTMAN, D. A. The measurement and motivation of atypical opinion in a certain group. *Amer. pol. sci. Rev.*, 1924, 19, 735-760.
- ALLPORT, F. H., & HARTMAN, D. A. The prediction of cultural change. In S. A. Rice (Ed.), *Methods in social science*. Chicago: Univer. Chicago Press, 1931. Pp. 307-350.
- ALLPORT, F. H., & SOLOMON, R. S. Lengths of conversations: A conformity situation analyzed by the telic continuum and J-curve hypothesis. *J. abnorm. soc. Psychol.*, 1939, 34, 419-464.
- ASCH, S. E. Effects of group pressures upon the modification and distortion of judgments. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (2nd ed.) New York: Holt, 1952. Pp. 2-11. (a)
- ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952. (b)
- CARTWRIGHT, D., & ZANDER, A. (Eds.) *Group dynamics*. Evanston: Row, Peterson, 1953.
- DOOB, L. W. *Social psychology: An analysis of human behavior*. New York: Holt, 1952.
- FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, 57, 271-292.
- FRENCH, J. R. P., JR. The disruption and cohesion of groups. *J. abnorm. soc. Psychol.*, 1941, 36, 361-377.
- HARE, P., BORCATT, E. F., & BALLS, R. F. (Eds.) *Small groups*. New York: Knopf, 1955.
- HYMAN, H. H. The psychology of status. *Arch. Psychol.*, N. Y., 1942, No. 269.
- KELLEY, H. H. Two functions of reference groups. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952. Pp. 410-414.
- KOFFKA, K. *Principles of gestalt psychology*. New York: Harcourt, 1935.
- LEWIN, K. *Principles of topological psychology*. New York: McGraw-Hill, 1936.
- LEWIN, K. Frontiers in group dynamics. *Hum. Relat.*, 1947, 1, 143-153.
- LEWIN, K. Field theory in social science. (Ed. by D. Cartwright) New York: Harper, 1951.
- NEWCOMB, T. M. *Social psychology*. New York: Dryden, 1950.
- SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, N. Y., 1935, No. 187.
- SHERIF, M., & SHERIF, CAROLYN W. *An outline of social psychology*. New York: Harper, 1956.
- THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups*. New York: Wiley, 1960.
- WOO C.-L. Conformity in custom behavior. Unpublished doctoral dissertation, Syracuse University, 1948.

## SOCIAL SCHEMAS

JAMES L. KUETHE

*Johns Hopkins University*

A person's perception of objects is often determined by the way in which he forms units. Many unit forming factors have been identified of which similarity, proximity, and set are a few (Koffka, 1935; Wertheimer, 1923). Consequences for interpersonal relations follow when people are seen as belonging together (Heider, 1958). The study of unit forming principles in social perception is required for an adequate psychology of interpersonal relations. Heider (1958) states, "A person may be seen in a cognitive unit with other persons because of kinship, nationality, or religion."

Unit forming principles in social perception can be regarded as social schemas or response sets to the extent that they function to structure ambiguous situations involving human objects.

When a person indicates that two objects "belong together" he has employed some schema or plan. If these objects are people or people-symbols, the schema employed may be considered, by definition, a social schema. When many people use the same schema in organizing a social response there is the implication that comparable experiences have produced the commonality of response. That the same response would be prepotent for many people would also be indicative of the pervasiveness of the tendency in the culture. Similarly, when situations result in low commonality of organization, it may be concluded that there are not shared experiences that result in the same predispositions for different people or that the shared predispositions do not tend to be prepotent.

If a person uses an idiosyncratic organization in a situation that most people organize with a high commonality schema, there are interesting implications. He may not have learned the modal response; but it is more likely (especially if he is from the same culture), that though he has learned the response, his personality dynamics prevent it from occurring. The prepotent response may

be rejected because of conscious or unconscious monitoring (Kuethe, 1960, 1961; Kuethe & Hulse, 1960). If the modal response aroused anxiety for a particular person, that response would not be prepotent for him.

De Soto and Kuethe (1958, 1959) have shown that subjects have and use schemas when they are asked about the relations existing between people. For example, given only the information that the people are acquainted, subjects think it rather probable that the people like each other. Some relations such as "liking" are assumed to be reciprocated and also to be transitive while other relations such as "confides in" are assumed to be reciprocated but not transitive. The subjective probabilities of these and other relations revealed the schemas people use when they think about the relations existing between people in minimal information situations. Expectations about interpersonal relations fell into three categories—symmetry, grouping, and ordering.

The investigation of social schemas often reveals principles to be found in common sense or naive psychology. The schemas are shared by all who belong to the culture; when thinking about them one tends to introspect and then, because he possesses the schema himself, say "of course" or "this is obvious." Heider (1958) states, "The veil of obviousness that makes so many insights of intuitive psychology invisible to our scientific eye has to be pierced." There is a tendency in psychology to ignore "obvious" principles of behavior, perhaps because of the desire to avoid things that appear trite or because their study does not seem "scientific." In doing so we run the risk of overlooking fundamental concepts, a failing not so prevalent in the more confident sciences.

The present investigation is concerned with those social schemas that determine whether or not objects are thought of as belonging together and degrees of "belongingness," that is, hierarchies of belonging when more than two



objects are involved. The central goals were to develop a technique for exploring social schemas and, using the technique, to investigate specific social schemas used by subjects in a variety of situations. In addition, the intention was to show that the social schemas used in organizing social responses are response sets in the sense that they can produce constant errors when subjects attempt to reconstruct situations that they have observed.

### EXPERIMENT I

#### Method

The subjects in this experiment were male undergraduates at the Johns Hopkins University; they performed the required tasks individually.

A piece of blue felt, 2 yards  $\times$  2.5 yards was

stretched on a wall of the experimental room. On each trial the subject was given two or more objects cut from yellow felt and was told to place them on the blue felt field in any manner he wished. The nap of the felt permitted the objects to cling wherever they were placed. The main advantage of the technique is that the objects may be placed anywhere on the field and with any orientation. When the objects are removed there is no mark left on the field that could influence future trials.

The complete lack of restraint on the nature of the response, permits the full operation of whatever schema is prepotent. The subject can either place the objects at random on the field or he can organize his response on the basis of some schema. Response sets of all varieties are permitted expression when people must perform in an ambiguous situation.

Each subject placed nine sets of objects on the field. The objects (except for the square, circle, and triangle) are shown in Figures 1, 2, and 3; they are:

1. Woman and child
2. Man and child
3. Three rectangles, each of different height
4. Man, woman, and child
5. Man, woman, and dog
6. Square, circle, and triangle
7. Man, woman, and two rectangles
8. Two women and two rectangles
9. Three men and three rectangles

The figures, except for the child and the dog, were between 7 and 10 inches tall.

The order in which the sets of objects were placed on the field was randomized for each subject to control for any influence that one set of objects might have on the reaction to another set. After the subject had placed the objects on the field, the experimenter recorded the relative placement of the objects and measured the distance between the objects. The experimenter then removed the objects from the field and gave the subject the next set of objects.

#### Results

It was immediately apparent that subjects responded to the task by giving organized responses; scattered or apparently random placement of the objects was, as will be shown, very rare.

*Woman-child and man-child object sets.* The woman-child and the man-child set were placed on the field in a vertical orientation (standing) by 94 of the 100 subjects. The woman and child were placed closer together than the man and child by 68 subjects, while 18 subjects placed the man and child closer together than the woman and child. This tendency was significant by a sign test,  $z =$

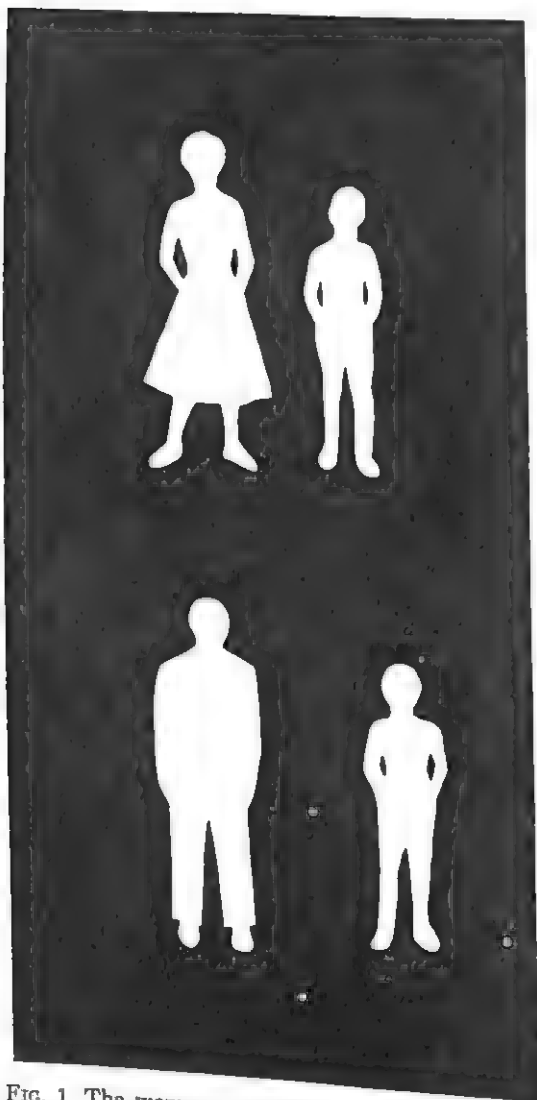


FIG. 1. The woman and child figures and the man and child figures.

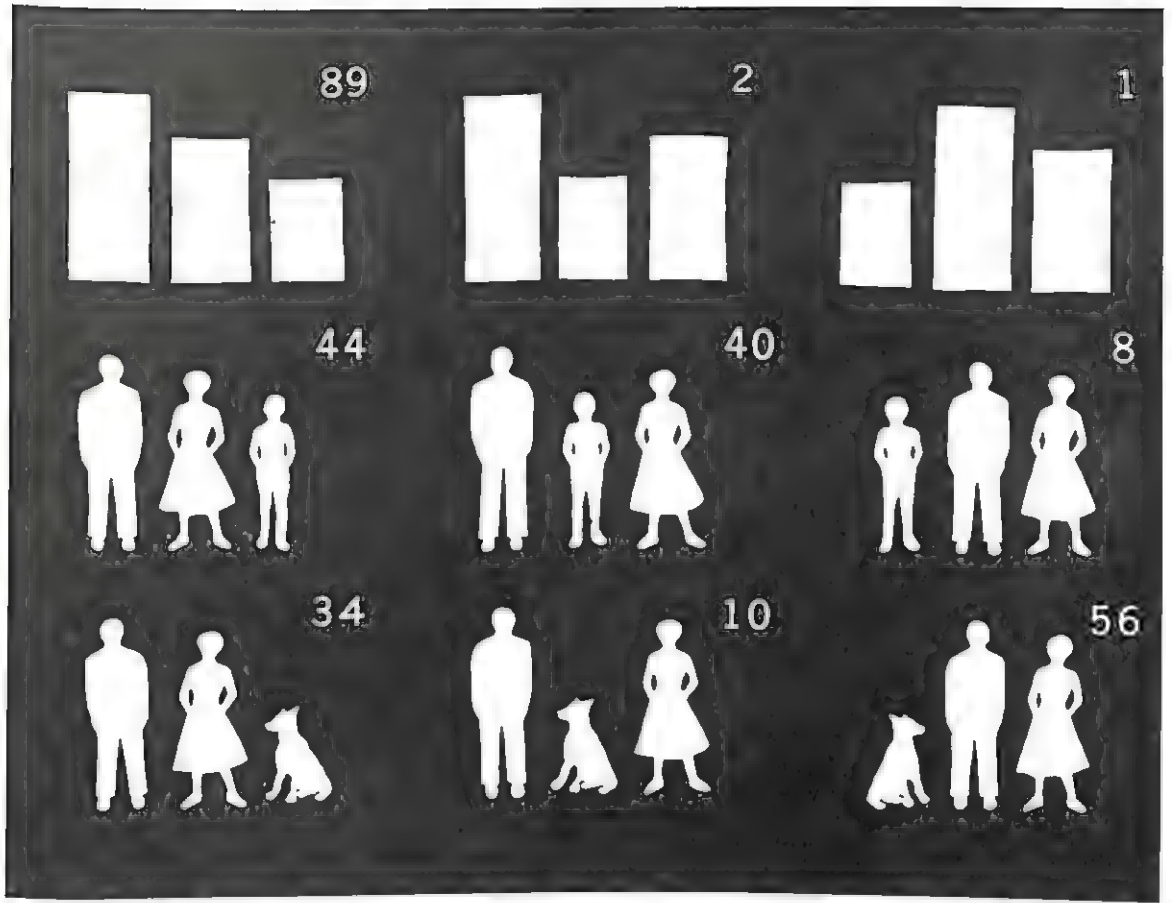


FIG. 2. Frequency with which 100 subjects used different orderings of the three rectangles; the man, woman, and child figures; and the man, woman, and dog figures.

5.30,  $p < .0001$ . There were 8 subjects that used an equal separation for the two sets of figures.

*Three object sets.* Reference to Figure 2 shows that the three rectangles of different sizes elicited a strong ordering schema based on height. Height ordering was used as the basis of organization by 89 of the 100 subjects. As may be seen in Figure 2, the other possible orderings were quite rare. This classification of responses was based on a pooling of the responses of subjects regardless of whether they ordered by height from left to right, right to left, or vertically. However, the strongest tendency was to order horizontally. Eight subjects used scattered arrangements that could not be classified.

The subjects could have consistently responded to the man-woman-child set by ordering them by height as was done with the rectangles. While an ordering by height was

the most popular, 44 subjects did this, the tendency to put the child between the man and woman was shown by 40 subjects (see Figure 2). It is quite interesting that only 8 of the 100 subjects put the figures in an ordering with the man between the woman and child. Eight subjects gave responses that can best be described as scattered. All of the 92 subjects who gave an organized response placed the figures in a horizontal row with the figures parallel and vertical as though standing on an imaginary base in the middle of the field. This is a basic schema in organizing human objects that is independent of schemas based on the specific content of the objects. It is the schemas based on specific content that determine the ordering of the objects in a set.

In contrast to the typical organization of the rectangles and the man-woman-child figures is the reaction of subjects to the man-



woman-dog figures. Of the 100 subjects who used this set, the order dog-man-woman was constructed by 56 subjects (see Figure 2). This set was included to permit the operation of a "man and his dog" schema, if subjects had such a schema. The order man-woman-dog was used by 34 subjects and may be an ordering by height schema. Only 10 subjects placed the dog between the man and woman; this is in contrast to the tendency to place the child there which was a quite popular mode of organization. That the dog was seldom placed between the people may reflect a schema whereby nonhuman objects tend not to be allowed to separate human figures; further evidence of this tendency will be seen later.

It is with the square-circle-triangle object set that we find the least evidence of a prepotent schema. The three possible orderings are used about equally often by the 76 sub-

jects who placed the objects in a line. The other 24 subjects who used this set placed the objects in unique patterns. Once again the tendency to use a horizontal placement was prepotent; it was used by 46 subjects while 30 subjects used a vertical arrangement. The failure of these objects to evoke a high commonality schema is revealed by the high proportion of idiosyncratic arrangements and the absence of a preferred ordering by those subjects who placed the objects in a row.

*Four object sets.* The tendency to place human figures together is quite strong. Figure 3 shows the popular response which is to place the rectangles outside of the people whether they are a man and a woman or are both women. This particular schema was used by 30 subjects with the man and woman figures and by 22 subjects with the two woman figures. However, many subjects placed the human figures side by side but

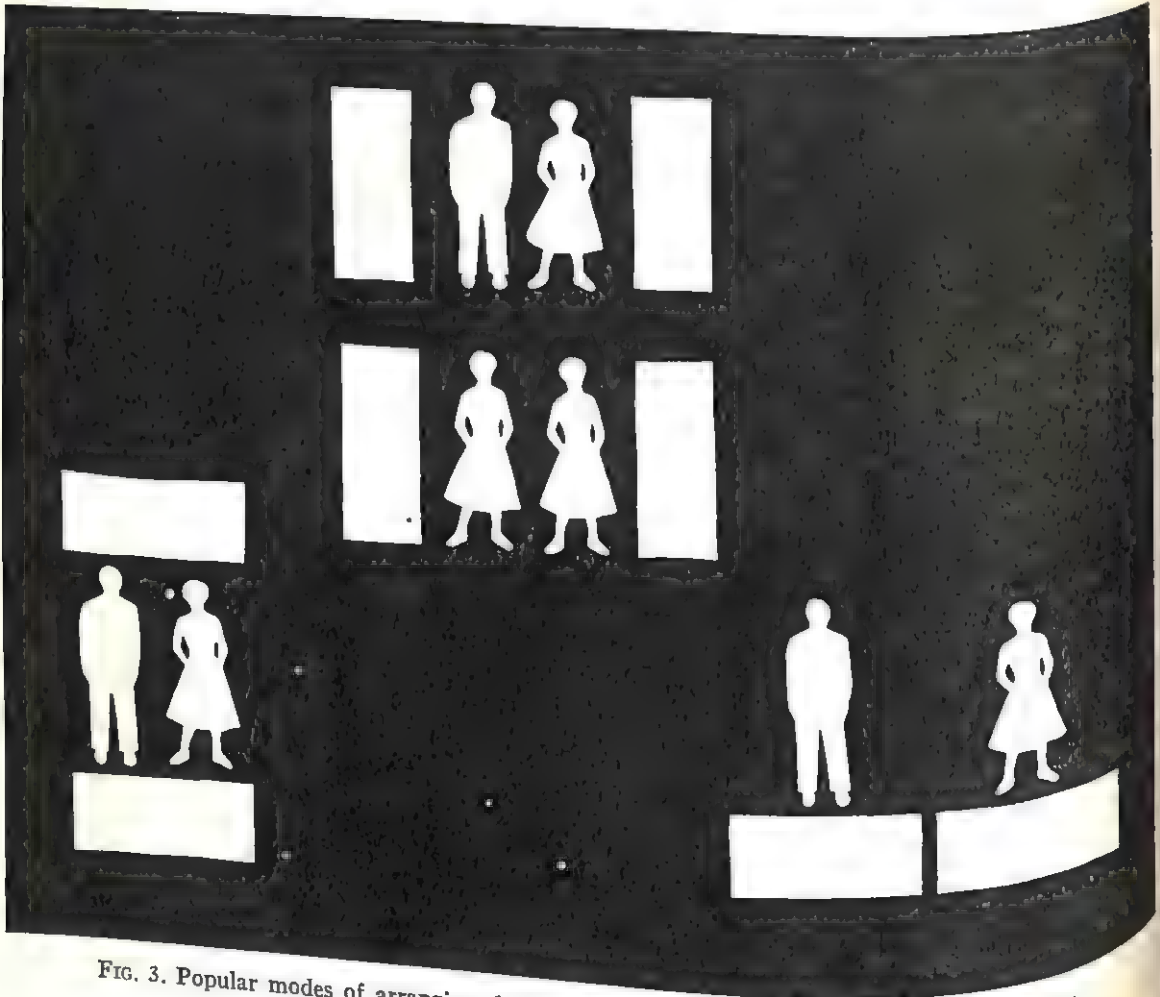


FIG. 3. Popular modes of arranging the man, woman, and two rectangles set of figures; and the two woman and two rectangles set of figures.

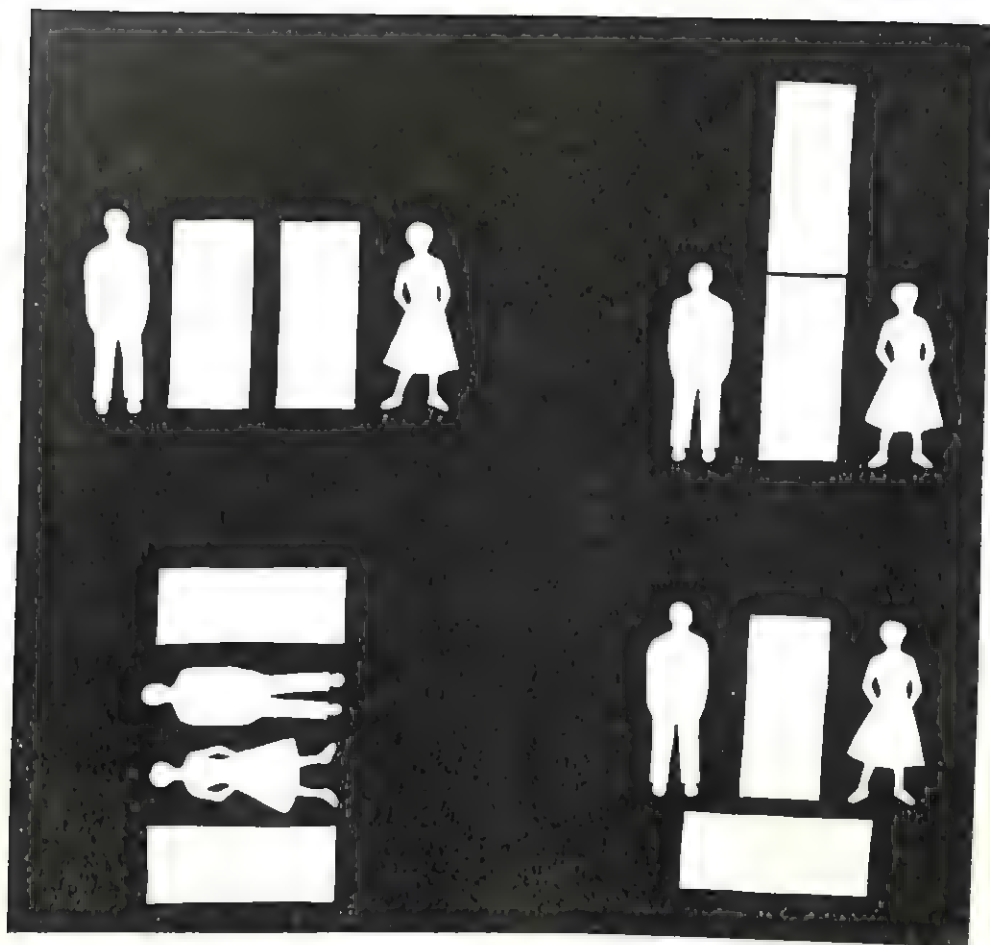


FIG. 4. Some idiosyncratic modes of arranging the man, woman, and two rectangles set of figures.

placed the rectangles elsewhere such as one above and the other below the people or both below the people. Once again the tendency to place human figures together without interposing nonhuman objects in manifested. This schema is especially interesting when it is considered that the two rectangles are identical. Their identity does not cause them to be grouped at the expense of grouping the human figures which, in the case of the man and woman figures, are not identical. This is more than a simple grouping of like figures, it is, rather, a social schema.

The subjects showed a stronger tendency to group the man and woman figures than the two woman figures. Out of 100 subjects, 78 grouped the man and woman to a greater degree than the rectangles,  $z = 5.50$ ,  $p < .0001$ . Only 58 of the subjects grouped the two women more than the rectangles; this

tendency was not significant,  $z = 1.50$ ,  $p > .10$ . This last analysis ignores the placement of the rectangles so long as they were not interposed between the people. The subjects were more willing to place one or both rectangles between the two women than they were to so separate the man and woman.

Two subjects placed the man and woman figures in a horizontal position with the man above, a rather creative response. This response and some other idiosyncratic responses are shown in Figure 4. Especially interesting are those responses in which rectangles are placed between human figures which are in marked contrast with the popular tendency to group human figures.

*Six object set.* The tendency to group human figures was again manifest. The three men were grouped to a greater degree than the three rectangles by 66 of the 100 subjects.



TABLE 1

## SUMMARY OF MODAL SCHEMAS IN EXPERIMENT 1

Stimulus set	Modal schema
Square-circle-triangle	Organized in horizontal or vertical plane, no preferred internal order.
Three rectangles of different height	Horizontal organization preferred, set ordered by height.
Man-woman-child	Usually ordered man-woman-child or man-child-woman. The order child-man-woman was rare.
Man-woman-dog	Preferred order was dog-man-woman. The order man-dog-woman was rare. Compare with man-woman-child in which the child was often placed between the man and the woman.
Man-woman-two rectangles	Usually the rectangles were not allowed to intervene between the man and woman.
Two women-two rectangles	Rectangles placed between two women often. Compare with man-woman-two rectangles.
Three men-three rectangles	The three men were grouped more often than were the three rectangles.

Typical patterns are shown in Figure 5. Only 6 subjects grouped the rectangles to a greater degree than the human figures. Organizations in which the human figures and rectangles were alternated were used by 20 subjects while essentially random displays were created by 8 subjects.

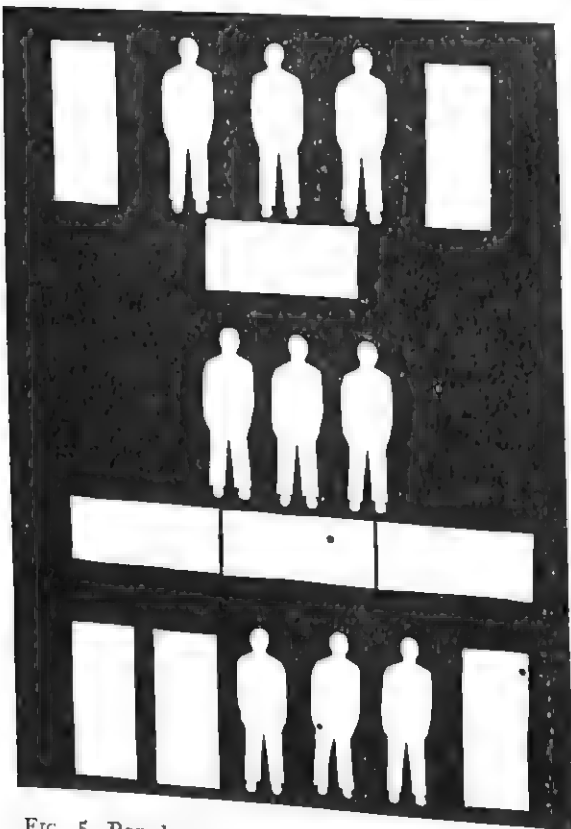


FIG. 5. Popular modes of arranging the three men and three rectangles set of figures.

## EXPERIMENT II

*Method*

The second investigation was concerned with whether or not social schemas function in reconstruction situations.

The two rectangles were placed 30 inches apart on the stimulus field. The first subject looked at this display for 5 seconds from a distance of 12 feet. The experimenter then took the objects down and gave them to the subject whose task it was to replace them exactly where they had been. Next, the experimenter measured the separation of the objects as placed by the subject. The man and woman figures were then placed on the field by the experimenter with a 30-inch separation and the procedure was repeated. The next subject started with the placement of the figures produced by the last subject. This sequence was continued until 30 subjects had reconstructed the displays produced by the subject that preceded them. For every other subject the order of presentation was reversed so that 15 subjects reconstructed the rectangle display first and 15 subjects reconstructed the man-woman display first. This technique of serial reproduction has been used to show the influence of social schemas by Bartlett (1932). The technique permits evaluation of a response bias over a total range rather than only at selected points. With this technique the basic schema can be seen when subjects no longer alter the configuration that is presented to them. Until stability is reached the responses of the subjects reflect the bias induced by the basic schema.

*Results*

The results of the second experiment are shown in Figure 6. From an initial separation of 30 inches, successive judgments brought the man and woman figures together until they were almost touching. As can be seen in Figure 6, once the male and female figures become very close together, the configuration is stable. Successive judgments do not bring the two rectangles together. Although there was individual variation, the separation of the rectangles after 30 successive judgments is of the same magnitude as the original 30-inch separation.

The schema that man and woman "belong together" induces errors of reconstruction. This is not a simple set to group like objects; the identical rectangles do not move together. As a check on the possibility that the technique introduced a bias, one group of 12 subjects replaced the man-woman display and the rectangles display after viewing them with a separation of 30 inches. In another group

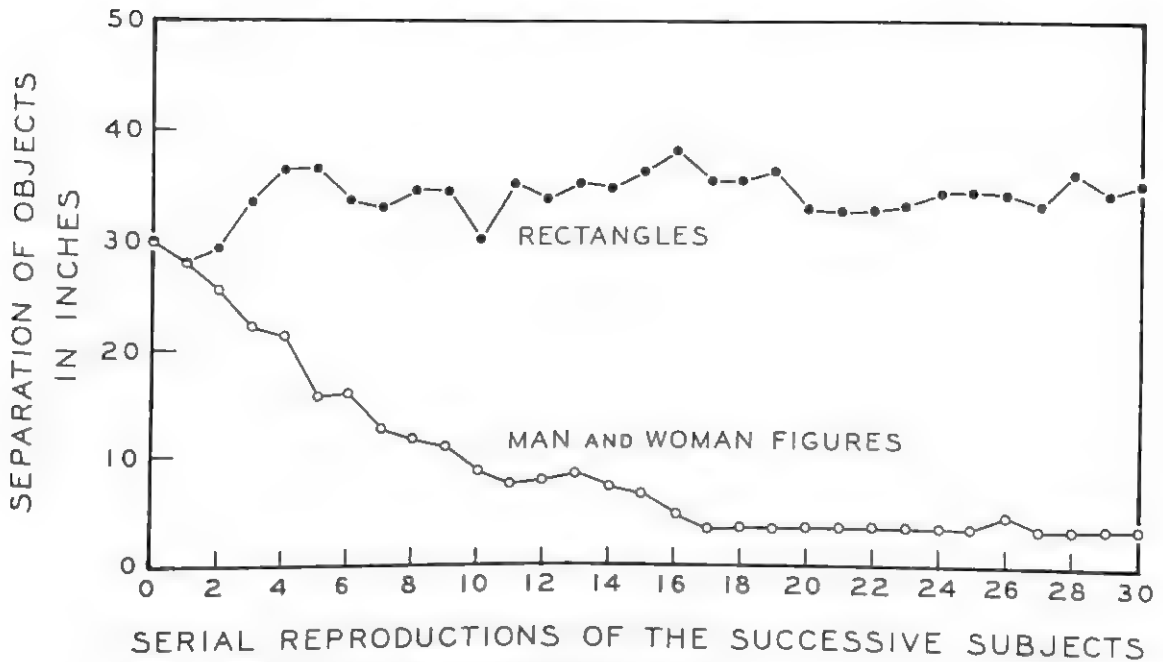


FIG. 6. Separation of the two rectangles and the man and woman figures following serial reconstruction by successive subjects.

of 12 subjects each subject replaced the two sets after viewing them with a 15-inch separation. The order in which the sets were reconstructed was different for 6 of the subjects in each group. The tendency to replace the man and woman figures closer together than the two rectangles was significant in both groups,  $p < .01$  by the Wilcoxon matched-pairs signed-ranks test.

#### DISCUSSION

When subjects are allowed to place sets of objects cut from felt on a field, their responses are organized even though the objects could have been placed with any orientation. Many arrangements had high commonality indicating that the subjects used the same schemas as a basis of response. The content of the objects in a set determined which schemas would be prepotent for subjects using that set.

After the basic tendency to organize and the tendency to use the horizontal has been noted, it is interesting to consider the relation between the content of the objects in a set and preferred arrangements of the set. There is a very strong social schema—people belong together. Human figures were grouped to a

greater degree than were nonhuman figures when all the figures were in the same set. Again, this is not a simple tendency to group like objects; the man and woman figures were grouped to a greater extent than were the identical rectangles. The same result occurred in a pilot study in which the man, the woman, and two circles were used.

The schema that people belong together and that nonhuman objects should not intervene has its parallel in the tendency of subjects to assume an affinity between people. Given only the information that the people are acquainted, subjects assume that the people like each other. Man's social nature is revealed by his readiness to assume that the relations existing between people are primarily positive and that interaction rather than isolation is the rule.

Social schemas vary in commonality with the nature of the human figures in a set. The male subjects in this investigation grouped the man and woman figures more often than they grouped the two woman figures. Common sense psychology would have predicted this result; most people share the schema that results in a strong association between man and woman. It may be that the schema



produces the verbal association just as it produces the spatial association measured here.

The three rectangles of different sizes were almost always arranged in a row in order of height, indicating that these objects tended to arouse a strong height ordering schema. The subjects could have arranged the man-woman-child figure set and the man-woman-dog figure set consistently by their height; the subjects did not do this. The content of these sets aroused specific social schemas that were prepotent over the height ordering schema. Of course, when subjects use a configuration with human figures that appears to be a simple height ordering, there is always the possibility that a social schema is being used which happens to be congruent with height ordering.

Although these investigations were designed to study high commonality schemas, the clinical significance of idiosyncratic responses should not be overlooked. A person's organization of the human figures may be identified as idiosyncratic once the typical responses of other people are known. For example, Figure 4 shows how a subject used the rectangles, one on top the other, to construct a high "wall" between the man and woman. This behavior attains significance because it contrasts with the strong tendency of most people to put the man and woman figures together. A response of this type might reflect a disturbance of the normal concept of the relation between male and female figures. Other responses made by subjects that could reflect disturbances of normal social thinking were grouping the non-human objects to a greater extent than human figures and placing the child next to the man on the side away from the woman.

There are many other combinations of figures besides the ones used here that would evoke social schemas shared by many people. The study of preferred modes of organization of unstructured situations reveals the schemas that typically operate. Identification of the basic social schemas will provide a foundation for the understanding of social perception and social thinking. Knowledge of high commonality principles of thinking about social objects will, in turn, permit full evaluation of idiosyncratic behavior.

## SUMMARY

Subjects placed sets of figures cut from felt on a felt board under conditions of free response. Most subjects responded to the task by giving organized responses; scattered or apparently random placement of the figures was rare.

The content of a set of figures determined the schema that was employed by the subject in organizing his response. A strong basic social schema results in the grouping of human figures to a greater extent than non-human figures. Several specific social schemas showed high commonality such as the tendency to place a child nearer to a woman than to a man and the tendency to place a dog nearer to a man than to a woman. The clinical significance of idiosyncratic responses was discussed.

A separate study showed that social schemas are social response sets in the sense that they can produce constant errors when subjects attempt to reconstruct situations that they have previously observed.

## REFERENCES

- BARTLETT, F. C. *Remembering: A study in experimental and social psychology*. New York: Macmillan, 1932.
- DE SOTO, C., & KUETHE, J. L. Perception of mathematical properties of interpersonal relations. *Percept. mot. Skills*, 1958, 8, 279-286.
- DE SOTO, C., & KUETHE, J. L. Subjective probabilities of interpersonal relationships. *J. abnorm. soc. Psychol.*, 1959, 59, 290-294.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- KOFFKA, K. *Principles of gestalt psychology*. New York: Harcourt, Brace, 1935.
- KUETHE, J. L. Acquiescent response set and the psychasthenia scale: An analysis via the ausgang experiment. *J. abnorm. soc. Psychol.*, 1960, 61, 319-322.
- KUETHE, J. L. The interaction of personality and muscle tension in producing agreement on commonality of verbal associations. *J. abnorm. soc. Psychol.*, 1961, 62, 696-697.
- KUETHE, J. L., & HULSE, S. H. Pessimism as a determinant of the tendency to claim undesirable symptoms on personality inventories. *Psychol. Rep.*, 1960, 7, 435-438.
- WERTHEIMER, M. Untersuchungen zur Lehre von der Gestalt. II. *Psychol. Forsch.*, 1923, 4, 301-350.

(Early publication received July 19, 1961)

# THE RELATIONSHIP OF INTERPERSONAL INVOLVEMENT AND AFFECTIVENESS OF CONTENT TO THE VERBAL COMMUNICATION OF SCHIZOPHRENIC PATIENTS<sup>1</sup>

STANLEY FELDSTEIN<sup>2</sup>

*Teachers College, Columbia University*

Despite the importance accorded schizophrenic communicative disturbances, few experimental efforts have been made to examine conditions which seem functionally related to them. Two such conditions suggested by the literature are interpersonal involvement, and affectiveness of communication content.

That interpersonal involvement affects schizophrenic communication is suggested by general interpersonal theory. It is perhaps most explicitly stated in a remark of Will (1954, p. xiv) which noted Sullivan's observation of the defensive aspect of schizophrenic language. Will proposed that, having experienced considerable anxiety in interpersonal relationships, the schizophrenic attempts to withdraw from them, either physically or by talking in a way that puzzles and irritates the listener. Few studies have manipulated the interpersonal variable experimentally and most of them have used some kind of written or drawn material as the interpersonal situation (Rabin & King, 1958, pp. 237-238).

<sup>1</sup> The article is based upon a dissertation submitted to Teachers College, Columbia University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This study was partially supported by USPHS Grant M-1052, "The Patients' Language as Expressive Behavior." The writer is indebted to George F. Mahl, principal investigator of the grant. The study was also partially supported by USPHS Grant M-4548.

Grateful acknowledgement is extended to members of the dissertation committee, Laurance F. Shaffer (Chairman), Joel R. Davitz, Elizabeth P. Hagen, and Edward J. Shoben, Jr. Grateful acknowledgement is also extended to the Psychology Service of Lyons Veterans Administration Hospital and the Psychology, Nursing, and Medical Illustration Services of the East Orange Veterans Administration Hospital for their unstinting support and cooperation.

<sup>2</sup> Now with the Research Department of the William Alanson White Institute, New York, New York, and the experimenter in this study.

Such studies tended to find that schizophrenics perform more poorly than normals. The two studies (Hirschman, 1953; Mirin, 1955) that attempted to explore the impact of live interpersonal situations on communication found that their schizophrenic subjects remained largely unaffected by increased interpersonal involvement. Their experimental conditions, however, introduced threats that did not arise out of the interpersonal situation itself and it seemed, therefore, unlikely that the results were solely the consequences of interpersonal involvement.

Theoretical and experimental investigations of how emotion affects the verbal communication of schizophrenics have almost always been couched in terms of its effect upon conceptualization (Feldstein, 1960, pp. 5-7). Jones (1957), too, although he examined the effect of emotion on the use of words by schizophrenics, was primarily interested in what his results revealed about the cognitive functioning of his subjects. To summarize such studies briefly, it may be said that affect and anxiety provoking stimuli have most often been found to disrupt and impair the thinking of the schizophrenic. It might be expected, therefore, that his language, too, would suffer.

Are schizophrenic communicative disorders semantic in character, or do they also involve expressive, or nonlexical disturbances? Schulze (1959), who tested the influence of erotic stimuli on several nonlexical attributes of schizophrenic speech, obtained no differences between his experimental and control groups. He pointed out, however, that his stimuli apparently failed to arouse his subjects sexually and, thus, he called into question the adequacy with which the hypotheses were tested.

The present study investigated the relationship of interpersonal involvement and affectiveness of content to certain lexical and non-



lexical attributes of spoken language that were considered significant aspects of effective communication. The three attributes selected were: the accuracy with which the perceived environment is reported; the relevance of what is reported; and the amount of disruption in the fluency, or word-to-word progression of speech. An *inaccurate element* was one that was obviously incorrect with respect to the stimulus material. An *irrelevant element* was a statement that involved a reference to situations not presented within the stimulus material. The *general Speech Disturbance Ratio (SDR)*, developed by Mahl (1959), was the measure of speech disruption used.

Inaccurate elements, irrelevant elements, and speech disturbances, then, were the three measures used in the study. Proportions of their occurrence were used in order to obtain measures independent of the size of the speech samples from which they came. The following hypotheses were based upon what appeared to be dominant implications of the literature.

A. Schizophrenics, compared with nonschizophrenics, communicate less effectively in a situation of increased interpersonal involvement than in a situation of decreased interpersonal involvement:

1. With increased interpersonal involvement schizophrenics, compared with nonschizophrenics, show an increased proportion of inaccurate elements.

2. With increased interpersonal involvement, schizophrenics, compared with nonschizophrenics, show an increased proportion of irrelevant elements.

3. With increased interpersonal involvement, schizophrenics, compared with nonschizophrenics, show a higher general SDR.

B. Schizophrenics, compared with nonschizophrenics, are less effective in their communication of affective content than in their communication of nonaffective content:

1. In the communication of affective content schizophrenics, compared with nonschizophrenics, show an increased proportion of inaccurate elements.

2. In the communication of affective content schizophrenics, compared with non-

schizophrenics, show an increased proportion of irrelevant elements.

3. In the communication of affective content schizophrenics, compared with nonschizophrenics, show a higher general SDR.

C. Regardless of the experimental condition the level of communicative performance is lower for schizophrenics than for nonschizophrenics:

1. Regardless of the experimental condition the proportion of inaccurate elements is higher for schizophrenics than for nonschizophrenics.

2. Regardless of the experimental condition the proportion of irrelevant elements is higher for schizophrenics than for nonschizophrenics.

3. Regardless of the experimental condition the general SDR is higher for schizophrenics than for nonschizophrenics.

Although expected directions were stated in the hypotheses, two-tailed tests of significance were used.

## METHOD

Thirty hospitalized, diagnosed, male schizophrenics formed the experimental sample; 30 male patients hospitalized for nonpsychiatric disorders, mainly from tuberculosis and orthopedic services, formed the control sample. Cases with organicity, psychosisurgery, and convulsive therapy were excluded. An unavoidable circumstance was that all the schizophrenic patients were receiving drug therapy. The mean age for each group was 35.5 years, with a range of 23-41 years for the schizophrenics and 20-50 years for the controls. None of the subjects had less than 8 or more than 13 years of schooling; the mean number of years for the schizophrenic sample was 10.3, for the control sample, 9.7. The difference between the means is nonsignificant. All the subjects were of average intelligence. Only patients who were willing to participate served as subjects. Only three patients refused to participate and were replaced.

The stimulus material, used to test the hypotheses concerned with the effect of content on communication, consisted of four picture sets, each set a series of five drawn pictures that told a story. An independent, nonpsychiatric sample judged two of the sets affective and two nonaffective. One of the affective sets depicted a story about a pregnant woman and a sick man, the other a story about a nurse and an angry patient. Of the two nonaffective sets, one depicted a story about a routine hospital day, the other a story about a man taking a book out of the library. Each picture was 5 × 7 inches and each set was mounted in sequence on a board 37 inches long and 7 inches high.

### *Procedure*

The interview was manipulated to explore the effects of increased (close) versus decreased (remote) interpersonal involvement. Half of each of the schizophrenic and nonschizophrenic samples participated in the close interpersonal situation; the other halves participated in the remote interpersonal situation. All the subjects were exposed to the affective and nonaffective material. To obviate the possibility of order effects, the order of the affective and nonaffective picture sets was changed from subject to subject.

Each subject was seen individually for the experimental interview, and each interview was fully recorded. The interview took place in an office in which only the subject and the experimenter were present. Two tape recorders were in the room and visible to the subject. One was used to record each interview, the other only to give instructions during the remote interpersonal situation. There had been no contact between any of the subjects and the experimenter previous to the experimental situation.

The close interpersonal situation was one in which two persons (the subject and the experimenter) interacted verbally and visually. The subject was brought, by a person other than the experimenter, into the interview room. The subject and the experimenter were introduced to each other. The subject was then asked to be seated. The experimenter attempted to involve the subject in a brief verbal interchange for the purpose of promoting a relatively close, albeit superficial, relationship. After several minutes, the experimenter began the instructions. In an effort to further decrease psychological distance between the participants the instructions made use of the first and second persons ("I" and "you").

The remote interpersonal situation was one in which two persons (the subject and the experimenter) were in each other's presence but did not interact verbally or visually. The subject was brought, by a person other than the experimenter, into the interview room and was asked by that person to be seated. The experimenter assumed the role of being primarily concerned with operating the tape recorders and did not look at the subject. The four boards holding the pictures were on a desk in the room ready for use by the subject. The instructions were issued by tape recorder soon after the subject was seated. Only the second person was used in the instructions in an attempt to increase psychological distance between the participants.

The instructions used in the close and remote situations are given by Feldstein (1960).

### *Scoring*

The scoring demanded simultaneous use of the tape recordings and verbatim transcripts of the experimental interviews. The smallest meaningful unit that could be considered an inaccurate element was a word. Proportions, therefore, were obtained by dividing the number of inaccurate elements in a speech sample by the total number of words in

the sample. The smallest meaningful unit that could be considered irrelevant was a sentence. Scoring of the irrelevant elements required, therefore, that the protocols of all the subjects be divided into sentences (Auld & White, 1956), and proportions were obtained by dividing the number of sentences in a speech sample into the number of irrelevant elements in the sample.

The scoring of speech disturbances, and the calculation of the general SDR have been described elsewhere (Mahl, 1956). The two component ratios of the general SDR, the Ah ratio (Mahl, 1959) and the non-Ah ratio, were also analyzed in the present study.

The experimenter did all the scoring. To obtain a measure of the reliability of scoring of the three measures used, samples of protocols were scored by two independent scorers. An interscorer reliability coefficient, obtained by correlating the average general SDRs for the complete speech samples of 20 subjects, was found to be .98. The same 20 cases were independently scored for inaccurate elements and another sample of 15 cases for irrelevant elements. Interscorer reliability for the proportions of inaccurate and irrelevant elements were .88 and .98, respectively.

The close and remote interpersonal situations were considered to be the two experimental conditions. Within each condition the subjects were given two trials, one affective, the other nonaffective. A modified form<sup>3</sup> of Edwards' (1950, pp. 288-296) analysis of repeated measurements was used to analyze the proportions of inaccurate elements, the general SDRs, and the Ah and non-Ah ratios. Arcsin transformations of all the proportions were used to normalize the distributions and equalize the variances. The transformations were multiplied by 100 before the analyses of variance were performed.

Because of the infrequency of their occurrence, the proportions of irrelevant elements were analyzed by means of a chi square corrected for continuity which compared the number of subjects who showed irrelevant elements within the two conditions and trials and within the experimental and control groups.

### RESULTS

Nonsignificant correlations were found between the proportions of inaccurate and irrelevant elements, between the proportions of irrelevant elements and the general SDRs, and between the proportions of inaccurate elements and the general SDRs. Because of this lack of relationship among the measures, they could not be assumed to be testing the three general hypotheses on effectiveness of communication and it was decided, therefore, to

<sup>3</sup> The writer is indebted to Jacob Cohen for his modification of Edwards' technique. The modification is described in Feldstein (1960).



TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE PROPORTIONS OF INACCURATE ELEMENTS USING ARCSIN TRANSFORMATIONS MULTIPLIED BY 100

	N	M	SD
Condition			
SC	15	31.40	10.08
SR	15	38.47	18.18
NC	15	36.67	18.86
NR	15	18.87	15.96
Trial			
SA	30	15.60	14.19
Sa	30	19.37	12.00
NA	30	11.10	11.30
Na	30	16.37	12.59

Note.—S=Schizophrenic; N=Non-schizophrenic; C=Close; R=Remote; A=Affective; a=Non-affective.

A comparison of the differences between the means of conditions SC and SR, and NC and NR yielded a  $t$  of 2.753 ( $p < .01$  level). Inspection reveals that the difference was not in the predicted direction. Interaction among Trials SA, Sa, NA, and Na was nonsignificant. Schizophrenic and non-schizophrenic performance across Conditions and Trials was not significantly different.

present the results in terms of the subhypotheses relevant to the two independent variables.

*Interpersonal involvement.* Hypothesis  $A_1$ —that schizophrenics, with increased interpersonal involvement, show a greater proportion of inaccurate elements than do non-schizophrenics—was not supported. Although the interaction between schizophrenic performance in the close and remote interpersonal situations and non-schizophrenic performance in the close and remote situations yielded an  $F$  of 7.69 ( $p < .01$  level), the difference was not in the predicted direction. It was the non-schizophrenics who, with increased interpersonal involvement, showed an increased proportion of inaccurate elements (see Table 1).

The protocols of 7 of the 15 schizophrenic subjects in the close situation contained irrelevant elements. None of the protocols of the schizophrenic subjects in the remote situation contained irrelevant elements. Nor were there irrelevant elements present in the protocols of any of the control subjects. A comparison of the subjects who obtained irrelevant elements in the two conditions yielded a chi square of 6.708, significant beyond the 1% level. Thus, Hypothesis  $A_2$ —that with increased interpersonal involvement schizophrenics, compared with non-schizophrenics,

show a higher proportion of irrelevant elements—was supported.

Hypothesis  $A_3$  stated that with increased interpersonal involvement schizophrenics, compared with non-schizophrenics show a higher general SDR. The interaction, however, between schizophrenic and non-schizophrenic general SDRs in the close and remote situations was not significant and failed to support the hypothesis. Similarly, analyses of the Ah and non-Ah ratios revealed no significant interaction between the groups and conditions (see Table 2).

*Affectiveness of content.* None of the hypotheses concerned with affectiveness of content ( $B_1, B_2, B_3$ ) were supported. Interactions between the schizophrenic and non-schizophrenic groups' proportions of inaccurate elements on the affective and non-affective trials were not significant. Similarly, four schizophrenic subjects on the affective and four on the non-affective trials showed irrelevant elements. The control subjects obtained no irrelevant elements on either trial. Interaction between the groups and trials, in terms of the general SDRs, yielded an  $F$  of 2.92, which is not significant. Analyses of the Ah and non-Ah ratios also revealed nonsignificant interaction between the groups and trials.

TABLE 2  
MEANS AND STANDARD DEVIATIONS OF THE GENERAL SDRs AND THE AH AND NON-AH RATIOS USING ARCSIN TRANSFORMATIONS MULTIPLIED BY 100

Condition	General		Ah		Non-Ah	
	M	SD	M	SD	M	SD
SC	128.93	27.81	96.13	27.17	80.53	23.65
SR	111.60	21.20	76.10	27.55	75.80	23.63
NC	101.40	23.38	65.67	28.41	69.80	23.81
NR	94.67	25.45	56.60	30.17	71.13	20.42
Trial						
SA	63.27	15.98	45.15	14.49	42.03	12.97
Sa	57.00	12.84	40.90	16.72	36.13	12.04
NA	49.63	13.81	30.40	14.38	37.10	12.61
Na	48.40	11.88	30.73	16.45	33.37	12.45

Note.—The N for each Condition is 15, for each Trial, 30. S=Schizophrenic; N=Non-schizophrenic; C=Close; R=Remote; A=Affective; a=Non-affective.

Interaction among the four Conditions and among the Trials yielded nonsignificant  $F$ s for the general SDRs and non-Ah ratios. Comparisons of schizophrenic and non-schizophrenic performance across Conditions and Trials yielded nonsignificant  $F$ s for the general SDRs and Ah ratios in terms of the Ah ratios an  $F$  of 11.60 ( $p < .01$  level).

*General performance levels.* It was hypothesized that schizophrenics obtain higher proportions of inaccurate elements ( $C_1$ ) and irrelevant elements ( $C_2$ ), and obtain higher general SDRs ( $C_3$ ) than do nonschizophrenics. A comparison of the proportions of inaccurate elements made by the experimental and control group yielded an  $F$  of 2.91 indicating that the general levels of the groups' performances were not significantly different. Thus, Hypothesis  $C_1$  was not supported. Hypotheses  $C_2$  and  $C_3$ , however, were supported. A comparison of the schizophrenics (7 out of 30) and nonschizophrenics (0 out of 30) whose protocols showed irrelevant elements resulted in a chi square of 5.700, significant at the 2% level. An  $F$  of 12.30 ( $p < .01$  level) indicates that the general SDRs of the schizophrenic sample were significantly different from those of the control sample over all conditions. Inspection of the means (Table 2) shows the difference to have been in the predicted direction. Further analyses showed the difference between the mean Ah ratios of the two groups ( $F = 11.60$ ) to be the major contributing factor to the difference between the general SDRs.

It should be noted that there were no significant interaction effects among the conditions and trials: the communication of emotional or nonemotional content was not particularly affected by the closeness or remoteness of the interpersonal situation.

*Additional findings.* In addition to the analyses pertinent to the hypotheses, the verbal productivity of the experimental and control groups was examined. Productivity was defined as the number of words spoken. Interaction among the groups and the conditions yielded an  $F$  of 3.83. A further test of the differences between the mean differences resulted in a  $t$  of 1.956 which cannot be considered significant. Nor was the interaction among the groups and trials found to be significant. Again, there was a nonsignificant difference between the general levels of schizophrenic and nonschizophrenic productivity. The analysis did indicate, however, that for all types of subjects the communication of affective content resulted in significantly greater productivity than the communication of nonaffective content ( $p < .01$

level), and the close interpersonal situation elicited greater productivity than the remote situation ( $p < .05$  level).

Both the experimental and control subjects responded with higher general SDRs ( $F = 6.51$ ) and higher non-Ah ratios ( $F = 9.53$ ) to the affective trial than to the nonaffective trial.

Correlations between the Ah and non-Ah ratios for the schizophrenics, nonschizophrenics, and total of 60 subjects over all conditions were found to be .18, -.09, and .11, respectively, all nonsignificant. Correlations were also obtained between the non-Ah ratios and productivity for the schizophrenics, nonschizophrenics, and the total number of subjects and were found to be .23 (nonsignificant), .47 ( $p < .01$  level), and .33 ( $p < .01$  level), respectively.

#### DISCUSSION

The most general implication of the results is that, in terms of the measures used, interpersonal involvement and affectiveness of content are not relevant parameters of the communicative disorder of schizophrenia. In the case of affectiveness of content, the implication is clearly supported. In the case of interpersonal involvement, the evidence is not so clear. The finding that schizophrenics introduced more irrelevant elements than the controls in the close situation qualifies the contention that interpersonal involvement is not a relevant parameter. It must be remembered, however, that less than half the 15 schizophrenics in the close situation made any irrelevant elements.

The results of the present investigation support the implication of the Hirschman (1953) and Mirin (1955) studies that the communication of schizophrenics compared to that of nonschizophrenics seems relatively unaffected by differing degrees of interpersonal involvement. The implication seems to derive a certain amount of validity from having been arrived at through a variety of conditions and measures. It is vital to remember, therefore, that in each of the three studies, the two persons who became interpersonally involved with each other were strangers. Whether the response of schizophrenics would be the same



to a relationship that existed over an extended period of time, such as that with a close friend, therapist, or "significant other," is a question that strongly merits further research. With regard to the effect of emotional content, the measures used in this study are not readily comparable to those used by previous studies concerned with schizophrenic thinking. The emphasis of the present investigation was upon communication rather than conceptualization.

Inasmuch as the three measures were found not to be correlated with each other, it may be useful to discuss the results in terms of each measure. Inaccurate elements were intended to measure distortions in the communication of the perceived environment. The measure is most closely related to that used by Aborn (1950), which described a tendency to respond in terms that in some way were inconsistent with the terms given in the stimulus material. An important difference between the two measures lay in the fact that, whereas Aborn's subjects were required to recognize correct descriptive terms, the subjects in the present study were asked to recall and freely communicate their perceptions. In spite of this difference, the results of the two investigations were similar. The combined evidence seems to suggest that the communication of the perceived environment is not more distorted in schizophrenics than it is in nonschizophrenics.

Irrelevant elements were intended to measure the communication of overinclusive thinking. The fact that only 7 of the 60 subjects made irrelevant elements makes it difficult to draw any stable generalizations.

The general SDR represents an analysis of the facility with which the language is spoken. Little effort has been made in past explorations of schizophrenic language disturbance to distinguish systematically between distortions in the execution and distortions in the meaning of the language. The results of this study indicate that the execution of speech is more disturbed in schizophrenics than in normals. The two component ratios of the general SDR, the Ah and non-Ah ratios, have been explored separately, and some ideas about their meaning have been

formulated. The non-Ah ratio was found by Kasl and Mahl (1958) to be positively related to the current anxiety level of normal individuals in the act of speaking. An hypothesis concerning the nature of the Ah ratio is that it is a measure of cautiousness.<sup>4</sup> Two tentative hypotheses suggested by the present results are that: the communication of affective content raises the current anxiety level of schizophrenics and normals similarly; and schizophrenics are more cautious than normals.

Of the additional findings, it might be said that the results concerning productivity failed to support Hirschman's (1953) findings that the general level productivity was higher for his controls than for his schizophrenic subjects. Also, the fact that productivity was found to be related to the non-Ah ratio for the controls hints, for the normal person at least, variations in the number of words spoken may be an indication of his current anxiety level.

The inferences underlying the present research were drawn from theoretical considerations that stress the importance of interaction that stress the importance of interpersonal interaction and affectivity to the schizophrenic communicative disorder. The fact that the inferences were not supported casts doubt upon the theoretical basis. It may be that schizophrenia is a mode of existence that inures the patient to the impact of interpersonal contact and emotional stimulation.

Several limitations of the study deserve attention. One is the lack of demonstrated cognitive equality of the picture sets used as the stimulus material. Another limitation concerns the type of patients used in the experimental sample. The subjects had to be both willing and able to communicate verbally. It may be that those patients who fulfilled the requirements were "too healthy" to show the expected communicative disturbances. In other words, it is not possible to state how representative the samples of the schizophrenics were of the populations they supposedly represent. Perhaps the most important limitation is the fact that the experimental subjects were undergoing chemical therapy at the time of the experiment. How

<sup>4</sup> G. F. Mahl, personal communication, 1960.

ever, the effect of tranquilizers in therapeutic doses upon the measures used, and particularly on verbal fluency in terms of the SDR, has never been tested. Even indirect evidence, involving investigations of the effects of tranquilizers upon other aspects of speech, is surprisingly scarce. Studies, however, such as those which demonstrated that meprobamate did not affect stuttering (Holiday, 1959; Maxwell & Paterson, 1958), and that of Hill (1958) which revealed that Sparine did not significantly affect word associations of schizophrenics, suggested that tranquilizers may in fact have little influence upon a measure of speech disturbances, or perhaps, upon language measures in general.

#### SUMMARY

It was hypothesized that schizophrenics communicate less effectively than nonschizophrenics: (a) in a close as compared to a remote interpersonal situation, (b) in the communication of affective as compared to nonaffective content, and (c) regardless of the experimental conditions. Inaccurate elements, irrelevant elements, and speech disturbances were the measures used to define effective communication.

The hypotheses were, for the most part, not supported. The schizophrenics did not differ in accuracy in the close and remote situations, although the controls were significantly less accurate in the close situation. The schizophrenics, however, made more irrelevant statements in the close situation, whereas the controls made none in either situation. The speech fluency of both groups was similar in both situations. In terms of the measures used, there was little difference between the communication of affective and nonaffective content by the experimental and control groups. Under all conditions, the accuracy of the groups was similar, whereas the proportions of irrelevant elements and speech disturbances were higher for the schizophrenics. The results raise questions about the relevance of interpersonal involvement and affectiveness of content to schizophrenic communicative disorder.

#### REFERENCES

- ABORN, M. A study of rigidity and distortion in normals and schizophrenics with controlled verbal material. Unpublished doctoral dissertation, Columbia University, Teachers College, 1950.
- AULD, F., JR., & WHITE, ALICE M. Rules for dividing interviews into sentences. *J. Psychol.*, 1956, 42, 273-281.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
- FELDSTEIN, S. The relationship of interpersonal involvement and affectiveness of content to the verbal communication of schizophrenic patients. Unpublished doctoral dissertation, Columbia University, Teachers College, 1960. (LC Mic. No. 61-856)
- HILL, H. J. Effects of Sparine and a total push program on the pathological functions of paranoid schizophrenics. Unpublished doctoral dissertation, Temple University, 1958. (LC Mic. No. 59-1565)
- HIRSCHMAN, W. A study of the communication of learned meaningful material by schizophrenics and nonpsychotics and its relationship to interpersonal involvement. Unpublished doctoral dissertation, Columbia University, Teachers College, 1953.
- HOLIDAY, AUDREY R. The effect of meprobamate on stuttering. *Northw. Med., Seattle*, 1959, 58, 837-841.
- JONES, R. A. *Some effects of emotion on the use of words by schizophrenics*. Washington, D. C.: Catholic Univ. Press, 1957.
- KASL, S. V., & MAHL, G. F. Experimentally induced anxiety and speech disturbances. Paper read at American Psychological Association, Washington, D. C., September 1958.
- MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, 53, 1-15.
- MAHL, G. F. Exploring emotional states by content analysis. In I. Pool (Ed.), *Trends in content analysis*. Urbana: Univer. Illinois Press, 1959. Pp. 83-130.
- MAXWELL, R. D. H., & PATERSON, J. W. Meprobamate in the treatment of stuttering. *Brit. Med. J.*, 1958, 1, 873-874.
- MIRIN, B. The formal aspects of schizophrenic verbal communication. *Genet. psychol. Monogr.*, 1955, 52, 149-190.
- RABIN, A. E., & KING, G. F. Psychological studies. In L. Bellak (Ed.), *Schizophrenia: A review of the syndrome*. New York: Logos, 1958. Pp. 216-278.
- SCHULZE, G. Speech fluency and speech disturbances in schizophrenics prior to and during exposure to an erotic stimulus. Paper read at American Psychological Association, Cincinnati, September 1959.
- WILL, O. A. Introduction. In H. S. Sullivan, *The psychiatric interview*. New York: Norton, 1954. Pp. ix-xxiii.

(Received January 3, 1961)



## THE POSITION OF HYSTERICIS AND DYSTHYMICS IN A TWO-DIMENSIONAL FRAMEWORK OF PERSONALITY DESCRIPTION

H. J. EYSENCK AND G. CLARIDGE<sup>1</sup>

*Institute of Psychiatry, Maudsley Hospital, University of London*

The present authors have always maintained that psychiatric classification in terms of diagnostic labels is not an adequate method of description (Claridge, 1960; Eysenck, 1947, 1952) and that much of the well known unreliability of diagnosis in psychiatry derives from the acceptance in this field of medical practices which are not well suited to the problems at issue (Eysenck, 1960b). Instead, Eysenck has proposed a description in terms of a dimensional framework derived from empirical studies, by means of appropriate statistical techniques—factor analysis and multiple discrimination analysis (Eysenck, 1957). He has also suggested that it was to be expected that groups of psychiatric subjects bearing the same diagnostic label would be found relatively close together in the  $n$ -dimensional framework resulting from the execution of such a scheme as that proposed; the degree of correspondence between categorical-diagnostic and continuous-dimensional framework would of course depend largely on (a) the reliability of psychiatric diagnosis (which could be increased by only choosing subjects where there was considerable agreement among psychiatrists as to diagnosis), and (b) the adequacy of the objective tests chosen to measure the subjects position in the dimensional framework.

In the field of neurotic disorders, the evidence seemed to suggest (a) that two main factors at least were required to account for the personality differences between subjects belonging to different psychiatric categories, viz., those of *neuroticism* and *extraversion*—

*introversion*; (b) that subjects labeled "hystericis" and "psychopaths" had high scores on neuroticism and extraversion, while subjects labeled "anxiety states," "reactive depressions," or phobic, compulsive, and obsessional patients had high scores on introversion and neuroticism. (This introverted neurotic group was called "dysthymics.") Mixed neurotics were supposed to be intermediate with respect to extraversion-introversion, but of course also high on neuroticism. This theory goes back to Janet and Jung in its main form, although the precise connotation given to the terms has almost certainly changed to some extent in the intervening years (cf. Eysenck, 1960d, for a historical survey of the development of the concept of extraversion-introversion, and for a review of the experimental literature).

Lately, a questionnaire has been published, called the Maudsley Personality Inventory or MPI (Eysenck, 1959), which purports to measure neuroticism and extraversion-introversion with sufficient reliability (both split-half and repeat are between .8 and .9) to serve as criterion scores when the more laborious measurements of these dimensions of personality by means of a battery of objective tests is not feasible. In normal populations these two scales are independent, or very nearly so, and much evidence has accumulated to show that predictions made on the basis of Eysenck's (1957) dynamic theory of personality can often be verified by having recourse to this questionnaire (Eysenck, 1960a). Use with neurotic groups, however, has been less successful. It has usually been found that the two scales do not retain their orthogonality but correlate together to an appreciable extent ( $r_{EN} = -.45$  approximately); that hystericis have lower neuroticism

<sup>1</sup> We are indebted to the United States Army, under whose grant, No. DA-91-508-EUC-311, this investigation was carried out. We wish to acknowledge the cooperation of J. McGhie, Commanding Officer at the Royal Victoria Hospital, Netley, and of R. L. Herrington, who assisted in connection with the determination of the sedation thresholds.

cism scores than do dysthymics; and that hysterics not only fail to be more extraverted than normals but may actually have slightly more introverted scores. (Psychopaths tend to behave as predicted, having high neuroticism and extraversion scores.) The results perturbed Sigal et al. (1958) to such an extent that they concluded: "The results suggest that either hysterics and dysthymics cannot be used in the described manner [i.e., as criterion groups for extraversion-introversion], or that the E and N scales do not measure introversion-extraversion and neuroticism, or that both statements are true" (p. 147). In a reply, Eysenck (1958) pointed out some errors in the arguments presented, but there is no doubt that the facts as given have been duplicated in several successive investigations (Eysenck, 1959) and that we are faced with a choice between two possibilities. Either we must give up the notion that hysterics are more extraverted than normals, as well as being more neurotic, or we must seek for some distorting factors in the MPI which account for the anomalous results in comparing normal and neurotic groups (Eysenck, 1958, p. 251). Clearly a recourse to experiment rather than to argument is called for.

The type of experiment required is one in which the three groups in question (normals, dysthymics, hysterics) are administered a battery of objective tests, differentiating between these three groups. The results could then be treated along either or both of the following lines:

A factor analysis could be performed which should give two factors identifiable as neuroticism and extraversion-introversion; factor scores would then be estimated for each subject, and these should discriminate the members of the three groups in such a way that hysterics should have the highest extraversion scores, dysthymics the lowest, while both groups should have higher neuroticism scores than the normals.

A multiple discriminant function (canonical variate) type of analysis could be performed which should give rise to two significant latent roots, identifiable as neuroticism and extraversion-introversion; canonical variate scores could then be obtained by using the latent

vectors as weights and multiplying them by the scores on the tests. These scores should then discriminate the members of the three groups in the same way as the factor scores.

A study fulfilling some of these requirements has been reported by Eysenck, Eysenck, and Claridge (1960), except that they concentrated on tests of extraversion almost exclusively, and did not obtain significant discrimination between neurotic and normal groups. Both a factorial analysis and an analysis of discriminance was performed, and on both hysterics emerged as the most extraverted and dysthymics as the most introverted group (the scores derived from the two analyses correlated to the extent of .81). The present study was carried out to give a more extensive and definite answer to the problem by attempting to use tests which would measure neuroticism as well as extraversion-introversion.

## METHOD

### *Subjects*

Neurotic subjects at the Royal Victoria Hospital were selected on admission where the psychiatrists in charge of the case could make a definite diagnosis of anxiety state or hysteria. One psychiatrist's diagnosis was sufficient to admit the patient to the experimental group. The subjects selected were not extreme or "pure" cases, but were chosen by the psychiatrist on the assumption that there would be a good consensus of opinion among psychiatrists regarding the diagnosis. Of the available cases not excluded for reasons given below, less than a third were excluded as presenting too mixed or complex a psychopathology to attempt a ready classification. Cases of immaturity were not included, and patients with evidence of brain damage, psychotic involvement, or drug addiction of any kind (including alcohol) were rejected. In each group there were 14 male and 2 female patients. The normal control group consisted of 16 volunteers, of whom 15 were male and 1 female. All were engaged on various duties in the hospital, including that of nursing orderly, clerk, storeman, and laboratory technician. The mean ages of the three groups were: dysthymics = 27.91, normals = 23.67, and hysterics = 23.78. The dysthymics were significantly older than the other two groups, with an  $F$  of 3.239. On weight and intelligence, as measured on Progressive Matrices, there were no significant differences.

### *Tests*

*Sedation threshold.* This was assessed in terms of the effect of sodium amytal on a simple task, consisting of doubling the digits read out to the



subject at intervals of 2 seconds from a tape recording of random digits over earphones, while he was receiving a continuous intravenous infusion of the drug at the rate of 0.1 g/min. Errors were recorded in blocks of five trials, and the threshold was taken as the point midway between the last two blocks with less than 50% errors and the first two blocks in which errors exceeded 50%. In the majority of cases these blocks were consecutive. The amount of drug administered at this point was determined from a chart relating blocks of trials to drug received, and this dosage was then corrected for the weight of the patient, giving the threshold in terms of mgm/Kg.<sup>2</sup>

*Maudsley Personality Inventory.* The two scales of the MPI were included in the experiment because their inclusion would, if anything, work against the hypothesis under investigation; furthermore, it was considered worth while discovering whether the neurotic groups in this study behaved similarly to those in previous ones.

*Five-choice serial reaction task.* The subject was required to press one of five keys set in front of five lights, according to which of the lights went on; his response extinguished the light and switched on another one. The order of the lights was random over a series of 50, except that no light appeared twice in succession. The score taken was the number of responses for each minute of the work period. Ten minutes of continuous performance was followed by a rest of 5 minutes and a further period of practice of 1 minute.

The usual pattern of performance shown by most subjects on this test (Venables, 1959) is a gradual decline in performance level during the first 5 minutes, then an increase in speed during the second 5 minutes of practice. Following the rest the usual reminiscence effect appears, in the form of an abrupt rise in performance level. The measures here taken are starting level and total number of errors.<sup>2</sup>

*Spiral aftereffects.* The subject is asked to fixate a rotating single-throw 180-degree spiral, and to indicate the duration of the aftereffect. Four trials were given, the spiral being rotated for one minute each time; trials were alternately clockwise and counterclockwise, with a rest of one minute between trials. The means of the four scores thus obtained were taken as a measure of the subject's performance on this task.

### Predictions

The predictions made follow from Eysenck's (1957, 1960a) theory, according to which cortical inhibition is hypothesized to be generated more quickly and more strongly, and dissipated more slowly in extraverts than in introverts; they are, in

brief, that extraverts would have *low* sedation thresholds, *high* scores on the E scale of the MPI, *low* starting level and *high* number of errors on the reaction time task, and *short* spiral aftereffects. Predictions with respect to neuroticism are much less obvious, except with regard to the N scale of the MPI; in view of the repeated observation that in neurotic groups N correlates negatively with E we might also expect E to have a negative loading with neuroticism in this population. Poor performance on sensory-motor tasks appears to characterize neurotics (Easterbrook, 1958; Eysenck, 1952; Eysenck, Granger, & Brengelmann, 1957) so we might expect the error score to load positively and the starting level score to load negatively on neuroticism. Predictions with respect to the sedation threshold cannot be made with any confidence. Shagass and his colleagues (as quoted by Claridge & Herrington, 1961, in their discussion of this problem) report high thresholds as characteristic of "anxiety," but this is a two-valued concept having loadings on both introversion and neuroticism. We have predicted high thresholds in introverts; this leaves the loading of the threshold on neuroticism indeterminate.

### RESULTS

These six sets of scores had been decided upon from the beginning as constituting the material for analysis. In addition, the Manifest Anxiety (MA) scale was administered, but not included in the analysis. Mean scores for the six tests and the MA scale are given in Table 1, together with SDs and *F* ratios. It will be seen that with the exception of the Five Choice Test Error Score all tests discriminate significantly; it will also be seen that scores on the E scale put the hysterics well to the introverted side of the normals, thus exaggerating the fault found by Sigal et al. (1958). Scores on the sedation threshold level, the spiral aftereffect, and the Error and Starting Level of the five choice test, however, put the hysterics on the *extraverted* side of the normal group. We find here the first signs that no single test can be relied upon to decide upon the correct position of groups such as these relative to each other, but that multiple determination is required. The MA scale, as expected, shows highest scores for the dysthymic group, and lowest scores for the normals; this was predicted (Eysenck, 1957) on the basis that the MA scale is essentially a measure of neuroticism with an admixture of introversion, i.e., that it is specifically aimed at the dysthymic group.

<sup>2</sup> For a detailed discussion of the rationale and method of use of the sedation threshold and five-choice serial reaction task, cf. Claridge and Herrington (1961), Claridge (1961), and Venables (1959).

TABLE 1  
MEAN SCORES OF DYSTHYMICS, NORMALS, AND HYSTERICIS

Test	Dysthymics	Normals	Hysterics	F
Sedation threshold	10.18 $\pm$ 1.608	7.86 $\pm$ 1.313	6.43 $\pm$ 1.774	21.837***
E scale	18.31 $\pm$ 8.169	31.50 $\pm$ 7.575	24.62 $\pm$ 8.366	10.120***
N scale	36.62 $\pm$ 7.631	19.12 $\pm$ 8.392	31.19 $\pm$ 11.092	14.354***
Spiral aftereffect	18.79 $\pm$ 5.093	11.41 $\pm$ 4.125	8.85 $\pm$ 3.167	22.576***
Five choice test, starting level	90.14 $\pm$ 15.412	82.34 $\pm$ 11.331	77.51 $\pm$ 12.499	3.530*
Errors	11.12 $\pm$ 10.705	8.63 $\pm$ 8.748	19.88 $\pm$ 20.524	2.563 <sup>a</sup>
M.I. scale	17.00 $\pm$ 3.774	6.25 $\pm$ 3.699	13.69 $\pm$ 5.860	21.903***

<sup>a</sup> ns = not significant at the  $p = .05$  level.

\* Significant at the  $p = .05$  level.

\*\* Significant at the  $p = .01$  level.

\*\*\* Significant at the  $p = .001$  level.

### Multiple discriminant function analysis.<sup>3</sup>

Table 2 shows the latent vectors and the latent roots extracted from our set of six tests: both latent roots are fully significant at the .001 level of  $p$ . The first variate accounts for 60.99% of the variance, the second one for 39.02%. Mean variate scores are given in Table 3, and the actual positions of the 16 subjects in each of the three groups plotted against the two variates are shown in Figure 1. Lines drawn in the body of the figure at values of 22 for Variate 1 and 12.2 for Variate 2 show that Variate 1 completely and without overlap discriminates between hysterics and dysthymics; this variate may therefore be identified with extraversion-introversion. Variate 2 discriminates slightly less well between normals and neurotics; three members of each group are misclassified. It must of course be

remembered that the criterion here is less satisfactory than in the case of the hysteric-dysthymic dichotomy, where psychiatric diagnosis of a clear-cut kind was obtained. The normal subjects in this experiment were normal only in the sense of not, at the time, being under psychiatric care; as Fraser (1947) and Shepherd, Fisher, Stein, and Kessel (1959) have shown, such groups nevertheless contain some 10% of fairly definite neurotics. In spite of our failure to obtain perfect discrimination, therefore, we may perhaps be justified in identifying Variate 2 with neuroticism.

The mean positions of the three groups are shown in Figure 1 in addition to the positions of the individual subjects, and it will be seen that hysterics lie on the opposite (extraverted) side of the normals, as compared with the dysthymics who lie on the introverted side. Furthermore, the hysterics appear if anything more neurotic than do the dysthymics. These results support the original theory regarding the respective positions of the three groups. It should be noted in this connection that our normal sample was conspicuously more extraverted than would be likely to be found in an unselected group; these Army volunteers

TABLE 2  
LATENT FACTORS AND LATENT ROOTS

	$\lambda_1$	$\lambda_2$
Sedation threshold	1.0000000	1.0000000
E score	-.0519939	.261961
N score	.0429769	-.354949
Spiral aftereffect	.505898	.158374
Starting level	.0657491	-.073380
Errors	.0156142	.0496781

$\lambda_1 = .732548$   $p < .001$   
 $\lambda_2 = .468676$   $p < .001$

TABLE 3  
MEAN VARIATE SCORES

	$\lambda_1$	$\lambda_2$
Dysthymics	26.60	8.67
Normals	18.37	14.59
Hysterics	16.50	5.69
	(Extraversion)	(Neuroticism)



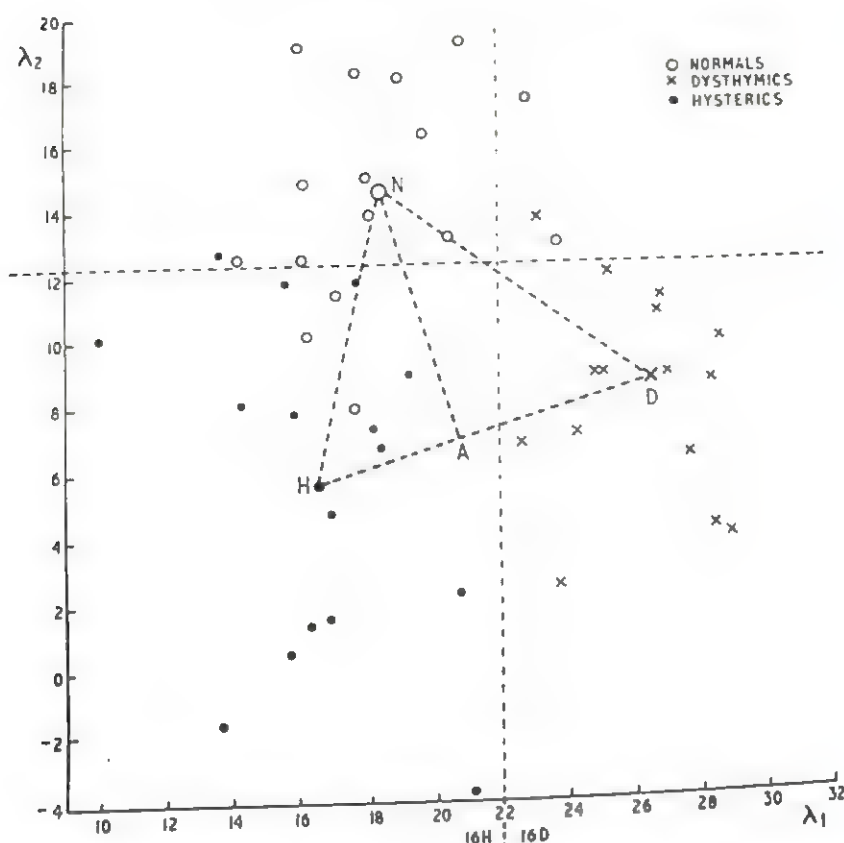


FIG. 1. Position of 16 normals, 16 dysthymics, and 16 hysterics in two-dimensional space as determined by analysis of discriminance.

had an E score of 31.50, as compared with the population norms of 24.91 given by Eysenck (1959). A similar tendency for Army volunteers to have exceptionally high extraversion scores had been noted in connection with a previous study (Eysenck, 1960a; Hildebrand, 1958). It is possible that this high degree of extraversion was due to the fact that the experimental subjects were volunteers; there is some evidence to suggest that extraverts are more likely to volunteer for experiments of this type. This might account for the fact that while the hysterics are more extraverted than the normals, the difference on this dimension between hysterics and normals is in fact much less than between normals and dysthymics.

An alternative hypothesis to account for this fact, as well as the rather curious finding that the hysterics emerged as more neurotic even than the dysthymics, may be derived from the fact that the canonical variates as extracted from the data depend to some extent

on the precise tests used, and are not likely to be collinear with the "true" variates which would be extracted from an infinite series of relevant tests. It is thus possible that the line H-D drawn in Figure 1 might be a better approximation to the "true" extraversion-introversion dimension, while the line N-A would then be an improved approximation to the "true" neuroticism dimension. The change would be minimal, and the number of misplacements would not be affected; yet hysterics and dysthymics would be equidistant from the normals, in opposite direction, along the extra-introversion dimension, and also equidistant from the normals, in the same direction along the neuroticism dimension. Further research along these lines, with a better selection of normals, and a greater number of tests, would be required to decide between these hypotheses. Whatever the final verdict, the fact remains that as predicted two highly significant latent roots have emerged from our analysis, thus indicating that two

TABLE 4  
FACTOR LOADINGS

	1	2	3	2'	3'
Sedation threshold	.81	-.27	.29	+.39	+.05
E	-.63	-.36	.33	+.48	+.11
N	.61	.19	-.40	-.44	+.07
Starting level	.37	.54	.73	+.29	-.86
Error	-.27	.86	-.15	-.62	-.62
Spiral	.72	-.13	-.14	-.04	+.19

dimensions are required to account for the differences between the three groups studied, and that in this two-dimensional space the three groups are disposed in the form of an (almost exactly) equilateral triangle. The hypothesis that hysterics and dysthymics are, respectively, extraverted and introverted neurotics has thus been confirmed by this analysis.

*Factor analysis.* Product-moment correlations were calculated between the six scores for the 48 subjects, and a Hotelling principal components factor analysis performed; all six factors were extracted, and the analysis was carried out with three digits retained after the decimal point.<sup>4</sup> Table 4 shows the unrotated factor loadings of the first three factors. Inspection of the plot of Factors 2 and 3 indicated clearly that a rotation in line with Thurstone's principle of simple structure would sort out Factor 3 as a doublet loading on nothing but the two scores derived from the reaction time test, and therefore probably simply an artifact. Accordingly the rotation was performed, and Factors 2' and 3' are shown in the table. Following a policy of minimum rotation, which seemed advisable as it gives least room for subjective judgment or arbitrary statistical rules, no further rotations were carried out, and Factors 1 and 2' are shown in Figure 2. Factor 1 can clearly

<sup>4</sup> There must be some doubt about the advisability of calculating product-moment correlations on a combined set of scores derived from three groups hypothetically situated at different points of a two-dimensional space. Strictly speaking this is not permissible if the hypothesis is in fact true, but the distributions are sufficiently normal, and the overlap sufficiently large, to make it not completely out of the question that the results might have a modicum of meaning.

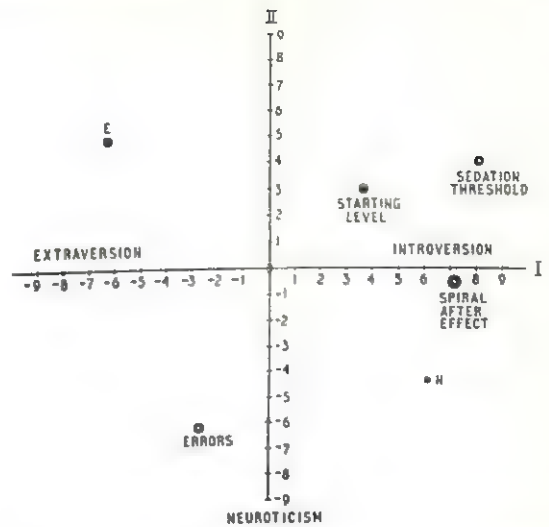


FIG. 2. Position of six tests in two-dimensional factor space.

be identified as introversion-extraversion, all the tests having the predicted loadings on it (sedation threshold, starting level, and spiral aftereffect are positive, E and errors negative). Factor 2 can be identified with neuroticism provided we are willing to accept the very tentative predictions made above.

The validity of both these interpretations can be tested by calculating factor scores for the members of our three groups; if the identification is correct, then these should fall into a pattern identical with that shown in Figure 1. Figure 3 shows the result of such an analysis, and the similarity with Figure 1 will be apparent. Such inspection is instructive, but a more quantitative estimate may be preferred. Factor 1 and Canonical Variate 1 correlate to the extent of .94; Factor 2 and Canonical Variate 2 correlate to the extent of .88. These correlations are high enough to give us some confidence in the identity of the two analyses, and in the accuracy of our factor identification. The cross-correlations, i.e., Factor 1 with Variate 2, and Factor 2 with Variate 1, are statistically insignificant.

As in Figure 1, so here also the baseline of the triangle produced by joining the mean positions of the three groups is not parallel with the horizontal axis, but inclined down towards the left; it follows that a slightly better discrimination might have been achieved by some further rotations. With the small



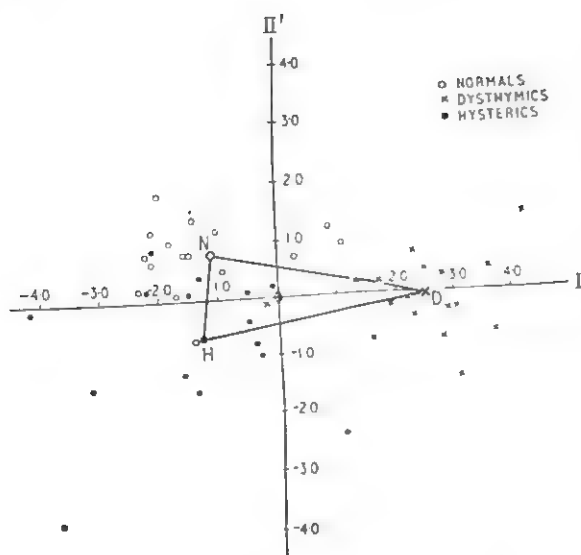


FIG. 3. Position of 16 normals, 16 dysthymics, and 16 hysterics in two-dimensional space as determined by factor analysis.

number of cases employed there seemed to be no point in further analysis, but it should be noted that here also, as before, the hysterics have higher scores on neuroticism than the dysthymics. As in the previous analysis, the hysterics are more extraverted than the normals, but only very slightly so; rotation to make the line linking the mean positions of dysthymics and hysterics collinear with the horizontal axis would increase this separation.

It might appear a task of supererogation to compare the results of these two methods of analysis (factor analysis and multiple discriminant function analysis) when it might be thought on a priori grounds that similarity would be the expected outcome. A mathematical and empirical investigation of the matter by Slater (1960) shows that this is not so, "Very little theoretical justification has been found here for expecting the vectors defined by a factor analysis to coincide with those defined by a discriminatory analysis of the same data; and the evidence examined has shown that they do not converge closely even under particularly favourable conditions." The data examined by Slater were those gathered by Hildebrand (1958) in one of the earliest factorial studies of the two-dimensional hypothesis under discussion (the data were collected in 1952); nothing can illustrate the improvement in the choice of tests for

the measurement of introversion-extraversion since those days better than the almost perfect congruence achieved in the present study, as compared with the failure to do so in the previous one. This improvement is likely to be due to the general theory advanced recently and purporting to give a rational basis to this dimension of personality (Eysenck, 1957), and conversely this success of the tests selected on the basis of the theory must give some support to the postulates of this theory.

#### DISCUSSION

The results of this analysis make it clear that hysterics and dysthymics, respectively, may indeed with advantage be used as criterion groups of the personality dimension of extraversion-introversion. It remains doubtful whether the hysteric group is as much more extraverted than the normal group, as is the dysthymic group more introverted than the normal group. It is likely that the proper counterpart of the dysthymic group is a combination of hysterics and psychopaths, with the psychopaths somewhat more extraverted than the hysterics; their omission from the group of supposedly extraverted neurotics may be responsible for the appearance of this doubt. No doubt remains about the degree of neuroticism of the hysteric group; in both analyses they had the highest scores on this dimension.

With respect to the MPI, the results leave little doubt, as had indeed been pointed out in the manual (Eysenck, 1959), that the E and N scales do not retain their independence when neurotic samples are being tested, or even normal samples with high neuroticism scores. Even under these conditions, dysthymics still obtain scores which put them at much more introverted part of the continuum than the hysterics, and the scales, therefore, retain some of their usefulness even when the level of neuroticism is high. However, comparisons involving both normal and neurotic groups become hazardous, and wrong conclusions may be drawn unless these essential cautions are borne in mind. Furthermore, spite of their high degree of neuroticism established by objective tests, hysterics have

lower scores on the N scale than do dysthymics; this point also requires caution in interpretation.

If these considerations only affected the MPI, it might be possible to brush the findings aside by refusing to use the test in question; this can hardly be said to resolve the problem, however, in view of the fact that in factorial studies considerable correlations have been found between the MPI on the one hand, and on the other, Cattell's second-order factors of extraversion and neuroticism, Guilford's various primary factor scales, the Taylor *MA* scale, and also some of the MMPI scales, notably the Hysteria, Psychopathy, and Psychasthenia scales (Eysenck, 1960a). What is true of the MPI, therefore, is *co ipso* likely to be true of these other scales too, and the Hildebrand (1958) study, which used a large number of different types of questionnaires, may be interpreted to support this position. Refusal to use the MPI, therefore, should entail logically the refusal to use any questionnaire, at least until this problem had been thoroughly investigated. Until this has been done, it would appear more reasonable to use these scales as before but bear in mind in the interpretation of results the nonlinearity of regression lines as the region of high neuroticism and high introversion is approached. Several hypotheses suggest themselves for experimental examination.

The hypothesis of response sets being responsible for the phenomenon cannot be ruled out, although some preliminary data speak against it (Eysenck, 1962). It is difficult to put such an hypothesis in a rigorous form, but this general field of investigation would almost certainly repay study.

The hypothesis of genuine interaction effects between introversion and neuroticism should not be disregarded. There may be a point in the conditionability, which according to theory characterizes the introvert, where fear responses are acquired at a rate and at a strength which exceeds the strength of the normal extinction processes; this point may delimit a region of positive feedback which could be responsible for the interaction. (Something of this kind is actually postulated by Wolpe, 1958, pp. 63-64, as an ex-

planation of the effectiveness of mild shock in learning neurotic responses.)

There may be a concentration on dysthymic symptoms in the construction of the N scale, and similar questionnaires, which leads to the omission of the main hysterical and psychopathic symptoms—possibly because these are difficult to elicit in a questionnaire (Eysenck, 1947). Milder forms of hysterical and psychopathic disorder might be more similar to those of dysthymics, thus making this problem urgent only at high levels of neuroticism. The *Hy* and *Pd* scales of the MMPI have not succeeded in overcoming this difficulty; in administering them to groups of hysterics, psychopaths, and dysthymics we have found no significant or even suggestive relationship with diagnosis, even when the E scale did show reasonable discrimination (Eysenck, 1962). With normals these scales also tend to work rather better (Eysenck, 1960a).

It has been shown that there is no unitary trait of *sociability*, but that of the items collected into that factor by Guilford (Factor S) approximately half correlate with extraversion, but not with neuroticism, while the others correlate with neuroticism (negatively), but not with extraversion (Eysenck, 1956). Thus there are two sorts of social shyness: introverted ("don't like being with people, but don't mind if I have to") and neurotic ("would like to be with people, but am afraid"). In view of the large part questions on sociability play in the E scale, it is possible that at high levels of introversion this distinction breaks down to some extent, thus causing the sudden break in linearity.

Perhaps the most likely reason for the correlation between introversion and neuroticism among high scoring subjects is one which can be deduced from the general theory of extraversion, according to which cortical inhibition is stronger in extraverts than in introverts. This principle applies to all sensory inputs, proprioceptive as well as exteroceptive; it seems likely therefore that the perception of the autonomic activity characteristic of emotion is equally subject to such inhibition. Now strong and lasting autonomic reactions are of course characteristic of neurotics (and quite generally of individuals with high scores on the N scale). We would expect therefore that



in comparing introverted and extraverted neurotics the autonomic reactions of the latter, while equally strong to begin with, would soon be subject to strong inhibitory forces and thus go down to a much lower level as compared with the reactions of introverts. Now the questionnaire items relating to neuroticism tend to refer in the main to long continued autonomic reactions or their consequences, and these in terms of our theory would be expected to be more marked in introverts than in extraverts. This hypothesis, like the others mentioned, is of course susceptible to experimental proof; it seems to involve no *ad hoc* assumptions which cannot be deduced from our general theory and it seems capable of resolving the problem raised by Sigal et al. (1958).

From a more fundamental point of view, it might be argued that the terms "extraversion" and "introversion" have been used rather misleadingly in two different senses. According to Eysenck's (1957) theory, certain innate physiological properties of the central nervous system (the excitation-inhibition balance) lie at the basis of observable extraverted and introverted behavior patterns. The term "extravert" may be applied (a) to a person in whom the excitation-inhibition balance is tilted in the direction of high inhibition and low excitation (the *constitutional* extravert), or (b) it may be used to refer to the person who is behaving in an extraverted manner (a *behavioral* extravert). As explained in detail elsewhere (Eysenck, 1960c) the theory posits a positive relationship between constitution and behavior, but this relationship is not likely to be perfect. Environmental differences clearly will play an important role in determining the way in which constitutional factors express themselves. Thus it is possible, although unlikely, that a *constitutional* extravert may turn out to be a *behavioral* introvert, or a dysthymic neurotic (Foulds, 1959; Foulds & Caine, 1958).

Certain factors may have contradictory effects on these two variables; thus old age probably alters the *constitutional* balance in the direction of greater *extraversion* and the *behavioral* balance in the direction of greater *introversion*. It is not impossible that neurosis,

whilst leaving the constitutional basis unaffected, shifts the behavioral basis in the direction of greater introversion. To take but one example, normal extraverts are more sociable than normal introverts. It is not inconceivable that the presence of severe neurotic symptoms (whether of a hysterical or dysthymic type) may interfere with social intercourse, thus shifting the behavioral balance towards greater introversion in both extraverts and introverts. Questionnaires such as the MPI are sensitive measures of *behavioral* extraversion; objective tests such as those used in this experiment are probably sensitive measures of *constitutional* extraversion. Factorial studies such as those of Claridge (1960), and the one reported here, show that constitutional and behavioral measures are not unrelated, but it would be a mistake not to distinguish in principle between them. The interesting question arises therefore whether hysteria and dysthymia are more closely related to constitutional or behavioral extraversion; the present results suggest the former. If true, this would mean that from the point of view of testing some of the experimental predictions from Eysenck's inhibition-satiation theory of personality, hysterics and dysthymics as criterion groups would be preferable to normal criterion groups selected on the basis of the MPI, although the latter test would come into its own in the testing of predictions relating the constitutional and behavioral aspects of personality to each other.

#### SUMMARY

The results are reported of testing 16 normals, 16 hysterics, and 16 dysthymics by means of objective laboratory tests and questionnaires, and analysing the scores by means of multiple discriminant analysis and factor analysis. Both methods give rise to two main principles of classification, or dimensions of personality, which can be identified as *extraversion-introversion* and *neuroticism*, respectively; high correlations are found between the respective methods of ordering the 45 subjects along the two continua. Hysterics were found to be extraverted and neurotic; dysthymics were found to be introverted and neurotic; perfect discrimination was achieved

between hysterics and dysthymics, and reasonable discrimination between neurotics and normals. Analysis of the questionnaire scores verified previous results in showing a departure from linearity of regression at high levels of introversion and neuroticism scores. The implications of these findings are discussed in relation to the use of hysterics and dysthymics as criterion groups for the study of personality dimensions.

## REFERENCES

- CLARIDGE, G. The excitation-inhibition balance in neurotics. In H. J. Eysenck (Ed.), *Experiments in personality*. Vol. 2. London: Routledge & Kegan Paul, 1960. Pp. 107-154.
- CLARIDGE, G. The effects of meprobamate on the performance of a five-choice serial reaction time task. 1961, in press.
- CLARIDGE, G., & HERRINGTON, R. N. Sedation threshold, personality, and the theory of neurosis. *J. ment. Sci.*, 1961, in press.
- EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1958, **66**, 183-201.
- EYSENCK, H. J. *Dimensions of personality*. London: Routledge & Kegan Paul, 1947.
- EYSENCK, H. J. *The scientific study of personality*. London: Routledge & Kegan Paul, 1952.
- EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, 1956, **50**, 113-140.
- EYSENCK, H. J. *The dynamics of anxiety and hysteria*. London: Routledge & Kegan Paul, 1957.
- EYSENCK, H. J. Hysterics and dysthymics as criterion groups in the study of introversion-extraversion: A reply. *J. abnorm. soc. Psychol.*, 1958, **57**, 250-252.
- EYSENCK, H. J. *Manual of the Maudsley Personality Inventory*. London: Univer. London Press, 1959.
- EYSENCK, H. J. *Experiments in personality*. London: Routledge & Kegan Paul, 1960. (a)
- EYSENCK, H. J. (Ed.) *Handbook of abnormal psychology*. London: Pitman, 1960. (b)
- EYSENCK, H. J. Levels of personality, constitutional factors, and social influences: An experimental approach. *Int. J. soc. Psychiat.*, 1960, **6**, 12-24. (c)
- EYSENCK, H. J. *The structure of human personality*. London: Methuen, 1960. (d)
- EYSENCK, H. J. Response set, authoritarianism and personality questionnaires. *Brit. J. soc. abnorm. Psychol.*, 1962, in press.
- EYSENCK, H. J., GRANGER, G., & BRENGELMANN, H. *Perceptual processes and mental illness*. London: Chapman & Hall, 1957.
- EYSENCK, S. B. G., EYSENCK, H. J., & CLARIDGE, G. Dimension of personality, psychiatric syndromes, and mathematical models. *J. ment. Sci.*, 1960, **106**, 581-589.
- FOULDS, G. A. The relative stability of personality measures compared with diagnostic measures. *J. ment. Sci.*, 1959, **105**, 783-787.
- FOULDS, G. A., & CAINE, T. M. Psychoneurotic symptom clusters, trait clusters and psychological tests. *J. ment. Sci.*, 1958, **104**, 722-731.
- FRASER, R. *The incidence of neurosis among factory workers*. London: HM Stationery Office, 1947.
- HILDEBRAND, H. P. A factorial study of introversion-extraversion. *Brit. J. Psychol.*, 1958, **49**, 1-12.
- SHEPHERD, M., FISHER, M., STEIN, L., & KESSEL, W. I. N. Psychiatric morbidity in an urban group practice. *Proc. Roy. Soc. Med.*, 1959, **52**, 269-274.
- SIGAL, J. J., STAR, K. H., & FRANKS, C. M. Hysterics and dysthymics as criterion groups in the study of introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 143-148.
- SLATER, P. Experiments in psychometrics. In H. J. Eysenck (Ed.), *Experiments in personality*. Vol. 2. London: Routledge & Kegan Paul, 1960. Pp. 247-299.
- VENABLES, P. H. Factors in the motor behavior of functional psychotics. *J. abnorm. soc. psychol.*, 1959, **58**, 153-156.
- WOLPE, J. *Psychotherapy by reciprocal inhibition*. Palo Alto: Stanford Univer. Press, 1958.

(Received January 4, 1961)



## A FURTHER STUDY OF THE RELATIONSHIP BETWEEN RESPONSE BIAS AND PERCEPTUAL DEFENSE<sup>1</sup>

MICHAEL J. GOLDSTEIN, SAMUEL HIMMELFARB, AND WALDA FEDER<sup>2</sup>

*University of California, Los Angeles*

Ever since the phenomenon of perceptual defense was suggested, researchers have attempted to explain it through principles which avoid the idea of a perceptual blocking mechanism (Eriksen, 1958; Goldiamond, 1958; Wiener & Schiller, 1960). The main argument used by all of these writers is that the indicator of perception typically involves the subject's articulation of anxiety linked words and neutral words. Since strong avoidance habits have previously become conditioned to the anxiety words, the probability of using an anxiety word as a response is lower than the probability of using a neutral word as a response (negative response bias for anxiety words), and this inequality in usage rates can produce differential chance congruencies in accuracy scores which could result in the perceptual defense effect.

The senior author has shown in a previous study (Goldstein, 1959) that a negative response bias for anxiety words can be obtained under blank stimulus conditions, and that this response bias is of sufficient magnitude to produce significant differences in pseudo-accuracy scores. However, in the same study it was shown that when a discriminative stimulus was presented, there was still a difference in accuracy between anxiety linked words and neutral words which could not be removed when these accuracy scores were corrected statistically for response bias.

Since statistical correction is always less satisfactory than correction through experimental procedures, the present report concerns itself with two studies in which an attempt was made to isolate experimentally the variance in accuracy scores due to response bias from the variance due to the presence

of a stimulus. The work of Blackwell (1952, 1953) has indicated that certain types of forced-choice methods provide recognition thresholds which are influenced less by the type of extraneous variance which we are trying to remove from the perceptual defense data than are any of the methods which employ free report. In the light of Blackwell's findings, it was decided that what was needed was a forced-choice indicator of perception which did not require verbalization of the two classes of words with different probabilities of usage. A technique which we believed would serve this purpose was one which resembled the spatial forced-choice method developed by Blackwell (1953).

The use of such an indicator for a word recognition study requires the simultaneous presentation of more than one stimulus word and a task set for the subject to locate one of the words by determining its spatial position in the group of stimuli (upper left, lower right, etc.). Note that the use of a spatial forced-choice indicator of this sort markedly changes the psychological nature of the recognition situation. It is no longer necessary for the subject to determine the presence or absence of one of a class of stimuli (anxiety arousing or neutral words); instead, he is told a stimulus will be presented and he must merely discriminate it from the others which will also be presented. Hopefully, this method should provide a measure of visual recognition uncontaminated by response bias.

### STUDY I

#### *Method*

*Subjects.* Subjects were 40 introductory psychology students at UCLA who were required to volunteer for 6 hours of experimentation as part of the course. Twenty males and 20 females were randomly assigned to 2 experimenters, 1 male, 1 female, each of whom ran 10 male and 10 female subjects.

*First experimental session.* In the first experimental session, the subject was administered a 117-item word

<sup>1</sup> This research was supported by a grant from the UCLA Faculty Research Committee.

<sup>2</sup> The authors would like to express their appreciation to Ira Nathanson for serving as the experimenter in the first experiment.

association test from which the anxiety words and neutral words were selected. The composition of the test and the method for selecting the eight words (four anxiety, four neutral) for each subject has been described in detail elsewhere (Goldstein, 1960). In brief, the eight words selected consisted of four pairs of words; one word of each pair was an anxiety word and the other a neutral word. The words in a pair were of the same frequency of usage, were the same length, and began with the same starting letter.

*Second experimental session.* In the second experimental session 24 hours later the subject was given four 3" x 5" cards (subject cards), with an anxiety word and its paired neutral word typed in capital letters on each card. On two of the cards the anxiety word appeared above the neutral word and on the other two cards the neutral word appeared above the anxiety word. The cards were designated Cards 1, 2, 3, and 4. The serial order of the cards as well as which two of the four anxiety words would appear above the neutral word was determined randomly for each subject. Each anxiety word was always paired with the same neutral word throughout the experiment.

There were eight stimulus cards: two cards for each of the four pairs of anxiety and neutral words. Each stimulus card had the two paired words centered and typed in capital letters on the same line, separated by  $\frac{3}{16}$ ". For each pair one of the two stimulus cards had the anxiety word to the left of the neutral word and the other card had the anxiety word to the right of the neutral word.

The following instructions were read to the subject by the experimenter:

This is a study dealing with perception. Here are four cards: Card 1, Card 2, Card 3, and Card 4. [The four subject cards were lined up on the table in front of the subject.] On each card two words are printed. These pairs of words will be presented in the machine [Gerbrands tachistoscope] one pair at a time. Before each trial I will tell you by saying the card number which pair of words will be presented and which one of the two words to look for. You are to look in the eyepiece and tell me whether the word I asked for appeared on the right or on the left. Don't tell me the word; just say "Right" or "Left." The words will be presented fairly rapidly so that at times it will be very difficult for you to see them and you will frequently have the feeling that you are just guessing. Don't be disturbed by this; just guess on every trial. Once again, before I expose each pair I will tell you which pair it is and which word to look for, then I will say "Ready" to alert you. After the exposure tell me whether the word I asked you for appeared on the right or on the left. Before we start take a minute to familiarize yourself with the words on the cards. Then we will try some practice cards so that you will become used to the machine.

After the subject felt that he was familiar with the words on the cards, he was given four practice trials before the main series of trials was run. The four practice trials consisted of one presentation of each of the four pairs of words. On two of the trials the subject was asked whether the anxiety word was on the right or left and on two trials the subject had to determine the position of the neutral word. On half of each of these two sets of two trials the word the subject was asked for was on the left, and on half the word was on the right.

*Experimental design.* The main body of trials consisted of 80 paired stimulus presentations, divided into five blocks of 16 exposures, with a rest period of 1 minute between blocks. Within each block of 16 trials the order of presentation was randomly determined with the restriction that each of the four pairs be presented four times. In addition, for every pair of words, the word the subject was to locate, and the actual position of that word was completely counterbalanced within each block of trials.

To summarize, the present experimental design involved a complex factorial design with two between-subjects effects (sex of the subject and sex of the experimenter) and three within-subjects effects (type of word the subject was to locate, spatial position of the word, and trials, or blocks of 16 exposures).

*Exposure time used.* The stimulus pairs were exposed in the tachistoscope at an exposure time which had been found in pilot studies to yield accuracy scores which were above chance (50% in the two-choice situation) for most subjects. If during the first two blocks of trials the subject was hitting below chance, the exposure time was increased by 25 milliseconds for the subsequent blocks of trials. If the subject was hitting too far above chance, the exposure time was decreased by 25 milliseconds. Thereafter no other adjustments of exposure time were made.

## Results

The subjects' responses were scored for the number of correct calls Left and Right for both anxiety and neutral words within each of the five blocks of 16 trials. This scoring procedure resulted in a  $2 \times 2 \times 5$  cell table for each subject. Within each cell the scores could range from 0-4, with a chance accuracy score per cell of 2. An analysis of variance was performed on the accuracy data for all 40 subjects. The source table consisted of five main effects and their respective interactions. The main effects were Sex of the subject, Sex of the experimenter, Type of word the subject was to locate (anxiety or neutral), Position of the word, and Trials. The grand mean for accuracy was found to be significantly



TABLE 1  
MEAN NUMBER OF WORDS CORRECTLY IDENTIFIED BY  
TYPE OF WORD ASKED FOR AND ACTUAL  
LOCATION OF THAT WORD

Actual position of the word	Subject is asked to locate	
	Anxiety word	Neutral word
Left	26.00	23.70
Right	23.60	25.60

greater than chance ( $F = 44.65, p < .001, df = 1/36$ ). None of the main effects was significant. There was no significant difference in accuracy for anxiety words when compared to neutral words; that is, *there was no perceptual defense effect*. However, type of word did interact significantly with position ( $F = 11.01, p < .001, df = 1/684$ ). The means for Word  $\times$  Position summed over the five blocks of trials are presented in Table 1. The anxiety words were recognized more accurately than the neutral words when they were on the left side of the stimulus card, while the reverse was true when the anxiety words were on the right side of the card. Note that it is not simply that one class of words was identified more accurately than the other class on only one side, but that the means are almost mirror images of each other for left and right sides. It should be further noted that there was no significant position effect, indicating that no absolute difference was derived from a stimulus word's presence on the right or on the left, unless the data are broken down by type of word.

Two other sources of variance were significant. They were Word  $\times$  Position  $\times$  Sex of

the subject ( $F = 13.14, p < .001, df = 1/684$ ) and Word  $\times$  Position  $\times$  Sex of the experimenter ( $F = 9.05, p < .005, df = 1/684$ ). Table 2 presents the mean accuracy scores for these two triple interactions. The means for the interaction of Sex of the experimenter with Word  $\times$  Position indicate that the trend noted above for Word  $\times$  Position was more pronounced when the male experimenter ran the experiment than when the female experimenter ran the experiment. In the case of the interaction of Word  $\times$  Position  $\times$  Sex of the subject the means indicate that the female subjects showed the Word  $\times$  Position effect, while the males did not. Examination of the data from the nonsignificant four-way interaction, Word  $\times$  Position  $\times$  Sex of the experimenter  $\times$  Sex of the subject, indicated that three groups showed the Word  $\times$  Position effect (male experimenter, male subject; male experimenter, female subject; female experimenter, female subject), but one group (female experimenter, male subject) showed an opposite effect; namely, that anxiety words were located more accurately than neutral words when they were on the right, and less accurately when they were on the left.

Further analysis of effects of word by position. Although the analysis of total accuracy scores revealed no evidence for the perceptual defense hypothesis, the presence of interactions between type of word asked for and its position in the visual field may have implications for this hypothesis. The critical issue is whether these interactions reflect perceptual behavior or whether they reflect some sort of bias in the use of the response words Left and Right.

TABLE 2  
INDIVIDUAL MEANS COMPRISING TRIPLE INTERACTIONS OF SEX OF THE EXPERIMENTER AND SEX OF THE SUBJECT WITH WORD BY POSITION

Word $\times$ Position $\times$ Sex of the experimenter				Word $\times$ Position $\times$ Sex of the subject			
Sex of the experimenter	Actual position of stimulus	Subject is asked to locate		Sex of the subject	Actual position of stimulus	Subject is asked to locate	
		Anxiety word	Neutral word			Anxiety word	Neutral word
M	Left	26.60	22.90	M	Left	24.60	24.20
M	Right	22.20	26.70	M	Right	24.90	24.10
F	Left	25.30	24.50	F	Left	27.30	23.20
F	Right	25.00	24.60	F	Right	22.30	27.20

Note.—Mean square error for testing both interactions was .84 with 684 df.

TABLE 3

MEAN NUMBER OF LEFT CALLS BROKEN DOWN  
ACCORDING TO ACTUAL POSITION  
OF STIMULUS

Actual position of the word	Subject is asked to locate		
	Anxiety word	Neutral word	Difference (A-N)
Left	26.00	23.70	2.30
Right	16.40	14.30	2.10
Left calls irrespective of position of stimulus	42.40	38.00	

To test the response bias hypothesis, we returned to the original protocols and counted the number of times the subject said "Left" when asked for the different classes of stimulus words. An analysis of variance was then done on these usage rate scores which revealed that every source of variance which was significant for accuracy was also significant for response bias (Word:  $F = 10.81$ ,  $p < .005$ ; Word  $\times$  Sex of the experimenter:  $F = 8.89$ ,  $p < .005$ ; Word  $\times$  Sex of the subject:  $F = 12.92$ ,  $p < .001$ ;  $df = 1/324$ ). No other sources of variance were significant. These significant  $F$ s indicate that whenever the subject was asked to locate an anxiety word he was more likely to use Left than Right as his response, while the reverse was true when the subject was asked to locate a neutral word. In an attempt to determine if these response tendencies were truly independent of the location of the stimulus, the data for number of Left calls were broken down by actual location of the stimulus. Table 3 below presents the mean number of Left calls for the interaction of Word  $\times$  Position. It can be seen in Table 3 that there were fewer Left calls for neutral words than for anxiety words, independent of the actual location of the stimulus. These data strongly support the hypothesis that the inequality in usage of Left and Right as recognition guesses is related to the type of word the subject is asked to detect and not to the word's position. However, to test this hypothesis of a response bias independent of the perceptual stimulus more rigorously, a second study was run.

## STUDY II

## Method

The second study was identical to the first except that no real words were flashed in the tachistoscope; instead, nonsense forms which looked like words were flashed. The subject was given the same instructions as in the first study, and no subject reported awareness of the true nature of the stimuli. The subjects were 10 male and 10 female students who were enrolled in a summer course in Psychological Aspects of Aging. The female experimenter was the same as in the first study, but a different male experimenter was used. Each experimenter ran 5 male and 5 female subjects.

## Results

The subjects' responses were scored for the number of Left calls made, and an analysis of variance was done on these data. The analysis failed to reveal any significant effects. However, inspection of the individual data suggested that rejection of the response bias hypothesis might be premature. Paralleling our first study, the majority of subjects showed an anxiety-Left, neutral-Right bias, while a minority of subjects showed an anxiety-Right, neutral-Left bias. The data for the 20 subjects in Study II are presented below in Table 4 broken down by type of bias (anxiety-Right or anxiety-Left). The data under the heading Left calls indicate the total number of times the subjects used Left as a response when asked to locate anxiety words and neutral words. The data under the pseudoaccuracy heading show the total number of correct stimulus detections the subjects made when the responses were scored according to the order of real stimulus presentations used in Study I.

The pseudoaccuracy column in Table 4 indicates that the inequality in Left calls for

TABLE 4  
RESPONSE BIAS AND PSEUDOACCURACY SCORES  
FOR BLANK STIMULUS STUDY BROKEN  
DOWN BY TYPE OF BIAS

	Left calls (Response bias)		Pseudoaccuracy scores	
			Left	Right
Right bias subjects ( $N = 6$ )	A	72	A	37
	N	162	N	78
Left bias subjects ( $N = 14$ )	A	341	A	172
	N	243	N	115



the two classes of words is sufficient to produce the same type of mirror image tables for the interaction of Word  $\times$  Position that was noted in the first study. Evidently, the unusually strong level of the response bias in the anxiety-Right group was sufficient to cancel out the response bias of the anxiety-Left group so that our analysis of variance in Study II did not yield a significant interaction of Word  $\times$  Position.

Although we cannot claim to have exactly replicated the response bias data obtained in Study I, we can conclude that it is possible to obtain this type of interaction between word and position in accuracy scores without flashing real words. This evidence, taken in conjunction with the evidence from the first study which indicated the independence of bias from actual position of the stimulus, strongly suggests that the interactions of word and position in the accuracy scores of the first study reflect the operation of a subtle form of response bias and does not support a perceptual blocking mechanism hypothesis.

#### DISCUSSION

The results of the forced-choice studies strongly support a response probability theory of perceptual defense. In previous studies where the stimuli were selected in an identical fashion, subjects were drawn from a highly similar population, and an indicator of perception which was sensitive to response bias was used (Goldstein, 1959, 1960), we have repeatedly found a perceptual defense effect. In the present study, an indicator of perception which does not permit response bias to affect accuracy scores was used and we found no perceptual defense effect. Although the forced-choice spatial location method is not free of its own peculiar type of response bias, the counterbalancing of position of stimulus and type of stimulus word asked for by the experimenter prevents the response bias of word by position from affecting the total accuracy score.

The conclusions drawn from the forced-choice studies regarding the locus of the perceptual defense effect do, however, conflict with earlier conclusions drawn by the senior author on the basis of another study (Goldstein, 1959). In that study, a significant

residual variance was found when accuracy scores were statistically prorated for response bias. The possibility of some sort of perceptual blocking mechanism could not have been ruled out on the basis of these statistically corrected data. How can we reconcile the apparently contradictory conclusions drawn from these two studies?

First, we can attempt a reconciliation by focusing upon the logic of the conclusion drawn from the statistically corrected data. It is conceivable that when this sort of correction for differential usage of the two classes of words (anxiety and neutral) is made, there is in fact, a true residual effect remaining in the data. However, this residual difference in accuracy between anxiety words and neutral words does not necessarily support, as we originally believed, the hypothesis of a perceptual blocking mechanism. There may be other types of response biases influencing the accuracy scores which can act independently of differential usage for the two classes of stimulus words: Data in perceptual defense studies are typically reported in terms of mean detection rates or mean thresholds for two contrasting classes of stimuli. Writers frequently overlook the fact that a class of stimuli is most often made up of a number of different stimuli. Thus, not only is it possible for the subject to show a response bias between classes of stimulus words, but it is also possible for him to show a within-word set bias resulting from the manner in which he distributes his guesses within a set. Since most perceptual defense experiments present each class of stimulus words with equal probability, a departure from this usage rate will affect accuracy scores for each set of words. Furthermore, words within a set are presented with equal frequency (.25, .25, .25, .25 for a four-word set) so that any departure in response rates from these probabilities will also be capable of affecting accuracy scores. If the subject distributes his responses for neutral words comparatively evenly within a set (.25, .25, .25, .25 for each word) while he distributes his responses within the anxiety set less symmetrically (.50, .30, .15, .05), then the probability of chance congruencies between the response and the stimulus presented is greater for the neutral words than for the

anxiety words. Such a within-set response bias could be partially independent of a between-set bias. The subject could use anxiety words 40% of the time and neutral words 60% of the time as recognition guesses; however, within each class of words a .25 call rate could exist for each word. This hypothesis is currently being experimentally tested in the UCLA laboratory with a design which permits the usage rates within a word set to vary while preventing any between-set bias.

Second, we can attempt to reconcile the conclusions drawn from the two sets of studies by focusing upon the logic of the forced-choice spatial location method. Although rigorous theory has never been conspicuous in discussions of the perceptual defense effect, most theoretical writings emphasize that the motivation for perceptual defense is anxiety reduction which can be achieved by the organism by avoiding the clear awareness of threatening stimuli. When the subject is free to describe his perception by verbalizing the word which he thought was presented in a brief tachistoscopic flash, then such a process can obviously operate. However, in the present study, the subject was told on each trial what stimuli were being flashed and no longer was there any ambiguity with respect to the presence of a particular stimulus in the immediate perceptual situation. Possibly, we have undermined the motivation for perceptual defense by insisting that the subject first *hear* the word from the experimenter before he *sees* it himself. The nature of the response bias (Word  $\times$  Position) found in the forced-choice studies indicates that verbalization of anxiety linked words by the experimenter is followed by a response pattern different from the one which follows verbalization of neutral words by the experimenter.

It is difficult to ascertain how seriously the above issue invalidates the forced-choice spatial location method in a test of the response probability theory of perceptual defense. If an investigator wishes to obtain some measure of perception uncontaminated by response bias, he can either statistically correct the data for response bias, or he can try to devise some experimental technique which prohibits response bias from operating. In the case of perceptual defense, we believe experi-

mental control to be superior to statistical control. However, no one method of controlling for response bias permits a crucial experiment since any one method may violate one of the conditions necessary for the occurrence of the perceptual defense effect. Only after a number of experiments which use different methods for controlling response bias have been completed, will there emerge a pattern of results that will indicate the necessity of postulating a perceptual blocking mechanism which inhibits the input of threatening stimuli.

#### SUMMARY

The present research was conducted for the purpose of determining if the perceptual defense effect is exclusively caused by the factor of response bias. In the first study, 20 male and 20 female subjects were given the perceptual task of identifying anxiety words and neutral words by specifying the word's location in space (left or right side of fixation point). The words were selected from each subject's word association protocol. The subjects were run by either a male or female experimenter. The results of the study were as follows:

1. By using a forced-choice spatial indicator of perception, which does not permit response bias to affect total accuracy scores, no perceptual defense effect was found.
2. A significant effect of Word  $\times$  Position was noted, which appears, upon closer analysis, to be due to a subtle form of response bias. When asked to locate anxiety words subjects responded "Left" significantly more than "Right." The reverse was true when the experimenter asked the subject to locate neutral words.

To determine whether the obtained response bias was independent of the presence of the stimuli, a second study was run. In the second study 10 male and 10 female subjects were used. The experimental situation was identical to that of the first study, except that no real stimulus words were flashed. However, the subject was instructed to expect real stimuli. The results of the second study were not identical to those of the first study with respect to patterns of response bias. However, inspection of the data from individual sub-



jects did reveal marked response biases of the type which could have produced the interaction between type of word and position found in the first study.

Two possibilities for the discrepancy between these results and an earlier study are suggested; first, that in the earlier study there may have been a within-word set response bias which may have affected accuracy scores even when the between-word set bias was removed. Second, it is suggested that the use of a forced-choice spatial indicator of perception may undermine the motivational basis for perceptual defense.

#### REFERENCES

BLACKWELL, H. R. The influence of data collection procedures upon psychophysical measurement of two sensory functions. *J. exp. Psychol.*, 1952, **44**, 306-315.

- BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *U. Mich. Engng. Res. Inst. Bull.*, 1953, No. 36.
- ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958. Pp. 169-227.
- GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373-411.
- GOLDSTEIN, M. J. Perceptual defense with and without a perceptual stimulus. *Amer. Psychologist*, 1959, **14**, 557. (Abstract)
- GOLDSTEIN, M. J. The effect of providing knowledge of results upon the perceptual defense effect. *Amer. Psychologist*, 1960, **15**, 582. (Abstract)
- WIENER, M., & SCHILLER, P. H. Subliminal perception or perception of partial cues. *J. abnormal soc. Psychol.*, 1960, **61**, 124-137.

(Received January 12, 1961)

## EFFECTS OF SOCIAL POWER ON ASPIRATION SETTING AND STRIVING<sup>1</sup>

ALVIN ZANDER AND THEODORE CURTIS

*Research Center for Group Dynamics, University of Michigan*

Early research on the level of aspiration provided a useful theory for explaining a person's selection of one goal from among a set of potential goals, his subsequent changes in this choice, and his evaluation of efforts to attain this goal (Lewin, Dembo, Festinger, & Sears, 1944). It was evident in those studies that selection of a particular level of aspiration and later changes in aspiration levels during a series of trials could usefully be conceived as coping behaviors, apparently intended to prevent a low evaluation of oneself. Later research demonstrated that a person's awareness of the scores obtained by others on the same task influenced his level of aspiration to the degree that he perceived the others to be like himself or to the degree that he desired to be like them (Chapman & Volkman, 1939; Festinger, 1942). More recent investigations have shown that a person can be directly influenced by others to place his level of aspiration at a particular point and to evaluate his performance in terms of the level proposed for him (Stotland, Thorley, Thomas, Cohen, & Zander, 1957; Zander, Natsoulas, & Thomas, 1960).

A quite separate set of studies has appeared, meanwhile, concerned with the power of one social agent over another, different bases of social power (Cartwright, 1959; French & Raven, 1959), and the efforts people make to evade or weaken pressures being exerted upon them by others (Zander, Cohen, & Stotland, 1957). Pertinent for the present purposes are the findings that a private belief is more likely to be influenced in certain types of social relationships (e.g., legitimate, expert, referent) than in the absence of these, or in a coercive relationship.

<sup>1</sup> The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

The convergence among the concepts and assumptions in these separate areas makes it possible to test hypotheses concerning the effects of social relationships upon coping behavior and the setting of a level of aspiration—hypotheses that are difficult to pose without a blending of level aspiration theory and theories of social power. Our interest was stimulated by findings from previous investigations (Stotland et al., 1957; Zander et al., 1960) which suggest that a person must "internalize" the aspirations others set for him if he is to work to attain the others' aspirations and to evaluate his efforts in terms of them. The purpose of the present research is to examine whether self-evaluations of performances and accompanying coping behaviors do in fact differ when persons are in a situation in which it is likely that they will accept the aspirations others set for them and when they are in a situation in which it is likely that they will reject such external inductions.

Several major assumptions are basic to the present approach. These are now stated with a general description of derivations one may make from them. Although these assumptions may apply either to persons who attain a level of performance proposed for them by others or to persons who fail to do so, this discussion is limited to individuals who fail to attain the level proposed by others. For convenience in notation, the letter *P* is used to refer to the person whose behavior we are predicting and the letter *O* to one or more persons in *P*'s life space.

*Assumption 1.* *P*'s evaluation of a given performance is a function of the size (and direction) of any discrepancy between his internal level of aspiration and his level of performance, his score.

We make a distinction between the overt level of aspiration and the internal level of



aspiration. The *overt* level is the value publicly expressed by *P* when requested to report the score he expects to attain on the next trial. The *internal* level of aspiration is a hypothesized value *P* privately uses as a standard when judging his performance as a success or a failure. The overt and internal levels may be exactly the same or may differ. The location of the internal level of aspiration, since it is a private matter, can be best inferred by an observer from the evaluation *P* assigns to a given performance, the higher his evaluation of the score the lower his internal level of aspiration. The first assumption then emphasizes that it is the internal level of aspiration which *P* employs in judging whether his performance is a success or a failure.

*Assumption 2.* *P*'s private cognitions are more effectively determined by social pressures originating in *O* when these pressures are based on referent power than when they are based on coercive power, assuming that the social pressures based on referent and coercive power are approximately equal in strength.

*Coercive* power of *O* over *P* is the ability of *O* to influence *P* because of *O*'s capacity to punish *P*. *Referent* power of *O* over *P* is the ability of *O* to influence *P* because of *P*'s desires to be similar to *O*. The theoretical reasons for Assumption 2 are presented by French and Raven (1959) and will not be described here.

In the present experiment three qualitatively contrasting conditions are presented to *Ps*: a coercive condition where *P* is pressed by *O* to achieve at least a given level of performance, with the threat of punishment if he does not do so; a referent condition where *P* has put before him the scores that attractive others have obtained on the task he is to perform; and, a control condition where *P* has no external standards of performance put before him.

The task *P* is to perform is one in which his actual level of achievement, or the level of performance set for him by *O*, can be expressed in a score; the higher the score the more difficult it is to achieve. On such a task each alternative score is a potential internal level of aspiration for *P*. Assuming that an

internal level of aspiration is a private cognition, it follows from the second assumption that *P* will place his internal level of aspiration closer to the level he perceives originating in *O* when he has a referent relationship with *O* than when he has a coercive relationship with *O*.

Referent and coercive social pressures of course can also have effects upon public behavior and thus upon the location of overt levels of aspiration. With respect to overt levels of aspiration, however, we assume that the two types of power are more nearly equal in their consequences than they are with respect to internal levels of aspiration and thus that overt levels of aspiration would be more alike between referent and coercive conditions than would internal levels of aspiration.

*Assumption 3.* When *P*'s score is discrepant from his internal level of aspiration, forces are generated in *P* to decrease this discrepancy. The larger the discrepancy the stronger these forces.

In a case where *P* performs below his internal level of aspiration on a given trial, the forces to decrease the discrepancy are assumed to evoke in anticipation of the next trial, either a lowered level of aspiration, an improved score, or both. A *P* in the control condition, for example, may be expected over a series of trials to reduce any gap between his scores and his levels of aspiration in either or both of these ways: first, perhaps, working until he reaches the ceiling of his ability and then lowering his level of aspiration when it is apparent to him that he cannot further raise his score. But a person placed under coercive or referent power, when exposed to levels of performance originating in *O* which are beyond the ceiling of his ability, has quite different problems in his efforts to reduce a discrepancy between his score and his internal level of aspiration.

In such a situation, a *P* in a referent relationship with *O* will experience strong forces from *O* to maintain his internal level of aspiration at a level beyond the upper limits of his performance. At the same time, as stated in Assumption 3, forces will be on *P* to reduce the discrepancy between his level of aspiration and his score. *P* may seek to decrease this

discrepancy by improving his future scores, which incidentally will also serve to make him more similar to *O*, but not by lowering his internal level of aspiration.

A *P* in a coercive relationship with an *O* who is attempting to get *P* to place his level of aspiration above the ceiling of his ability will not experience strong forces on his internal level of aspiration. In a series of trials, therefore, a *P* in a coercive relationship with *O* will lower his internal levels of aspiration when this is necessary to decrease a discrepancy between his level of aspiration and his scores, and will have no strong tendency to improve his scores.

Where the levels of performance originating in *O* are higher than *P* can attain despite his full efforts to do so, it follows that the average discrepancy between his internal levels of aspiration and his scores is likely to be greater in a referent relationship than in a coercive one. Given this situation, a *P* in a referent relationship will evaluate his performance lower than a *P* in a coercive one.

In effect, we propose that a *P* in a referent relationship will set his internal levels of aspiration in accord with those originating in *O* and will work hard to attain those levels, whereas a *P* in a coercive relationship will resist influence attempts to raise his internal levels of aspiration above his scores and will not strive to improve his performance. A referent relationship thus generates acceptance and a coercive relationship generates resistance to social pressures originating in *O*. We expect to find evidence of these contrasting tendencies toward *O* accompanying the above mentioned decisions by *P*. In a referent situation, for example, *P* should be attracted toward *O*, desire to have continued relationships with him, and state that *O*'s influence attempts are acceptable to him. In a coercive relationship, in contrast, *P* should be little attracted to *O*, desire to avoid further association with him, and discredit the effectiveness of *O*'s influence attempts upon him.

The performance of a task requires the use of one or more abilities. We assume that there is a ceiling to *P*'s abilities in his task beyond which he cannot perform even though he might aspire and try to do so. When *P*

cannot lower his aspiration or improve his performance enough to decrease the discrepancy between the two, the forces on *P* to reduce the discrepancy may generate tendencies to discredit the validity of his score and tendencies to discredit the centrality of the ability involved in the score, thereby reducing the importance of the discrepancy or decreasing in fantasy the amount of the discrepancy. From the prediction that the discrepancy will be greater in a referent than in a coercive situation, it follows that discrediting the validity of the score and the centrality of the ability is more likely to appear when *P* is in a referent relationship than when he is in a coercive one.

Lastly, the various forces and pressures on *P* may generate emotional tension. *Emotional tension* is the result in *P* of forces acting upon him in opposing directions. Indicators of emotional tension are reports of worry, feelings of muscular tension, and the like. We have seen in the above paragraphs that the nature of the opposing forces differ for persons in a coercive situation and for those in a referent one, but it cannot be said with confidence which situation will create the greater emotional tension. We may expect, at least, that control subjects are less likely to develop emotional tensions than participants in either of the two experimental treatments.

The contrasts anticipated in the behavior of a *P* in a referent situation and a *P* in a coercive situation should be further evident by comparing results from these experimental treatments with the results from the control condition. Given that control subjects are subject to almost no social pressures, we expect that the results from the control condition ordinarily will fall between the two extremes represented by the accepting referent subjects and the resisting coerced subjects, insofar as reactions to *O* are concerned. In other respects the behavior of control subjects should replicate the usual findings from "classical" studies on the individual's level of aspiration (Lewin et al., 1944).

#### METHOD

Participants in this experiment were brought into the laboratory one at a time. They were told



that they were a member of a team, would meet their mates later, and that their team would compete with teams from other schools. The "test" administered during the experiment was said to be preparatory to this competition.

### *Subjects*

The subjects were 60 high school senior boys in a suburban high school. Each boy met with the experimenter in a small examination room of the school health services.

### *Task*

The experimenter introduced himself as an employee of the University Testing Center (a fictitious organization) and explained that he was working on a nationwide program to test how much "small muscle control" high school students have. This ability was described as a skill in accurate muscular movements as well as a competence to judge how much control one has over these movements. The judgment aspect of the ability was included in order that the experimenters requests for overt levels of aspiration before each trial would appear reasonable.

The nature of the experimental task was briefly explained to the subjects and they were asked to make 15 practice shots (5 shots equal one trial). The task was the Rotter (1942) Level of Aspiration Board. The board, built to the exact specifications provided by Rotter, is 4 feet long, 2 inches deep, and 4 inches wide. A groove 1.25 inches wide and .5 inch deep extends the length of the wider surface. A .75 inch steel ball is propelled down this groove with a short cue stick. The object is to make the ball stop as close as possible to a Number 10 placed beside the groove. Numbers from 1 to 9 are placed .75 inch apart, decreasing in amount in both directions from the Number 10. The number reached by the ball is the score obtained by the subject for that shot regardless of whether it is beyond or below the 10. The higher the score, the better the subject has performed. Small indentations in the groove, next to each number, slow the rolling of the ball and steady it beside one or another number. During the administration of the test the board was raised 1.5 inches at the end distant from the subject.

This task was chosen because it is interesting, quickly learned, and requires little native ability. Rotter (1942) states that there is little improvement due to learning. Our experience has shown that more careful effort can produce a better score. Rotter has provided norms of typical scores which made it possible for us to set standards for the subjects which were beyond the ability of most persons and which gradually rose over a series of trials so that the standards from 0 consistently remained above their ability as the subjects improved with practice.

When the practice trials were finished the experimenter described in some detail why this test is a

valid measuring device and went on to explain the importance of small muscle control. He described changes in our modern world which make this ability increasingly important.

As we come to depend more and more upon delicate and complicated machinery, more jobs open up which require small muscle control. A jet pilot, for example, needs the ability to a high degree. He must be able to adjust controls quite exactly, making his muscles do what he wants them to do. He must be able to judge how much error he is likely to make, even when he tries his hardest to be exact. . . . And closer to home: as plants and factories become more automated, using more complicated and delicate machinery, such as electronic computers to control production, more and more important and responsible jobs on both a management and employee level will appear and have to be filled. . . . We are testing high school students now so that we can estimate how many men there will be in a few years who have enough ability to fit into such jobs.

It was explained that the test today would help the subject obtain information about himself relevant to the possibility of choosing certain careers.

In order further to increase interest it was stated that the teams would be asked to take additional tests on another day, during which they would be competing against other schools in this state.

The teams will gather in the state capital for a large tournament. Each school will have several 15-man teams. The best team will receive a prize and will be declared the winner.

Publicity about this program was being withheld, it was said, until the state tournament was about to be held.

Different instructions now followed depending upon whether the subject was assigned to the coercive, referent, or control condition.

### *Coercive Condition*

The coercive condition demanded that subjects be led to believe that potential teammates had the ability to punish them. The punishment in this instance was being expelled from the team and not being allowed to participate in the state tournament.

Subjects were told that at this school 3 or 4 extra boys had been chosen for each team. This meant that in order to bring the teams down to 15 members, the ceiling required by the rules, 3 or 4 boys must be eliminated from each team. It was said that the job of elimination was being given to a committee of 5 potential teammates whose names had been chosen from a hat. The names of the boys on this committee, subjects were told, would not be made public. The committee had met, had had the test explained to them, had not taken the test themselves, but had decided just about what score a person ought to be able to achieve on each trial. Subjects were told that they would be given the scores that the committee expected of them trial-

by-trial and that their own scores would be sent to the committee for judging. No exact criterion score was stated which would determine whether a subject would remain on the team or be expelled. Thus no subject could be wholly certain whether his level of achievement was making him subject to punishment.

### *Referent Condition*

To generate a referent relationship it was necessary to create in the subjects a desire to be like *O*.

In this condition participants were told that a committee of boys had been appointed to be co-captains of the teams from this school. These boys were chosen, it was said, by athletic coaches and teachers of English who, respectively, could judge the boys on motor control and judgment ability. The boys were also selected on the basis of their ability in leadership. The chosen co-captains were known to be liked and respected by students in this school, it was added, but the members' names were being kept a secret for the present. The results shown in the last row of Table 4 indicate that the subjects desired to be similar to the members of the committee.

"This committee has met and has taken this test. I'll show you their scores," the experimenter said, adding "the committee has asked us to send them your scores in order that they may have some idea how well this school's team is likely to do." The subjects were told that they were members of the team and would remain on it no matter what their scores were on this day.

### *Control Condition*

In this situation no social pressures were generated for the subjects. They were told that they were members of the team but no mention was made of requirements placed on them by others or of how well others had done on the test. It was said that a committee of five boys had been chosen by random selection and were to act as co-captains of the teams. The experimenter added that he was going to send the subject's scores to the committee in order that the committee might have some idea how well the teams from this school were likely to perform. Subjects were led to believe that they were permanent members of the teams.

### *Public-Private Treatments*

It was assumed in the original planning of this experiment that results would differ when comparing the reactions of subjects in a situation where *O* would know *P*'s scores and where *O* would not know *P*'s scores (Raven, 1959; Stotland & Zander, 1958).

In order to generate a public condition, the subjects were informed that the scores for one half of their trials would be reported to *O*. To create a

private condition subjects were told that the other half of their scores would not be reported and that these trials were for the subject's practice and information. Appropriate differences in method of recording the subject's scores and in reporting to the subjects the standards originating in *O* were developed for the two treatments. Within each influence type, half of the subjects had the private condition first and the other half had the public condition first, thus balancing the effects of sequence in experiencing the two conditions.

The public and private treatments were not different in their effects in any important respects. It is probable that the manipulation was not successful in creating contrasting public and private conditions, either because it incorrectly depended upon the ability of subjects to change their "set" from public to private while most of their cues were the same in both conditions, or because the presence of the experimenter was sufficient audience to make both situations public ones.

### *Administration of Experiment*

The procedure following the above instructions was similar for the two experimental groups. The entire series of scores originating in *O* (the same for each subject) was posted on a large sheet, one score for each trial. Before each trial, the subject was asked to estimate what his score would be for the next trial. This overt level of aspiration was posted, as well as the actual score obtained by the subject for that trial, in such a manner that discrepancies among the three entries were obvious to the subject. The levels of performance originating in *O* were in all cases above the subject's probable best score. Control subjects followed the same procedure, except that no scores from *O* were mentioned.

### *Measures*

After the first seven trials and after the second seven, subjects were asked to fill out a questionnaire. The questionnaire was introduced as an attempt to obtain *P*'s reactions to the experience so that the Testing Center could improve their tests and so that the center could better understand the nature of small muscle control. The instrument contained of small muscle control. The instrument contained Likert-type rating scale queries (seven-point scales) about self-evaluation, motivation, validity of the tests, demands of the *O*, desire to remain on the team, affective reactions to the situation, and emotional tension.

Following the second questionnaire the experiment was declared to be completed. The purposes, deceptions, and the reasons for these were explained and all questions from the subject were answered.

### *Analysis*

The results from this experiment were suitable for study by analysis of variance, using the Type



TABLE 1  
MEAN SCORES AND EVALUATIONS OF SELF

Query	Control	Coercive	Referent	<i>t</i> of differences		
	<i>M</i> (A)	<i>M</i> (B)	<i>M</i> (C)	A-B	A-C	B-C
Subjects scores on practice trials	26.0	26.1	26.6	<i>ns</i>	<i>ns</i>	<i>ns</i>
Standards originating in <i>O</i> (via the experimenter)	—	39.4	39.4	—	—	—
Subjects overt levels of aspiration	29.8	33.6	34.4	4.00***	4.89***	<i>ns</i>
Subjects performance scores	28.0	27.8	30.5	<i>ns</i>	2.77**	2.74**
Difference:						
Overt levels-performance	1.8	5.8	3.9	4.00***	2.14**	1.83
Subjects satisfaction with performance <sup>a</sup>	3.4	3.2	2.5	<i>ns</i>	2.64**	2.05**

<sup>a</sup> Query: How satisfied or dissatisfied do you feel with your performance on the test game-board? (Very dissatisfied—Very satisfied)

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

III method proposed by Lindquist (1953). These analyses made it evident that statistically significant effects in the interaction of the experimental conditions with public and private treatments, and with the order of administration of the latter, were so few and irrelevant to the theory that they may be ignored. Results from public and private conditions, regardless of their sequence are, therefore, pooled in the next section.

The *t* tests comparing the differences between the means in the referent, coercive, and control conditions were computed by using the error term in the within-subjects sum of squares as the error term when calculating the significance of the differences between means, as suggested by Lindquist (1953, p. 243). Differences between means for the separate experimental conditions were tested for significance only when the *F* tests for main effects were significant. Values of *p* in the tables refer to one-tailed *t* tests.

## RESULTS

### Ability on Practice Trials

On the three practice trials, before instructions to create different experimental conditions were given subjects by the experimenter, the average scores of the 20 subjects assigned to each of the experimental treatments were almost exactly the same. We can be confident then that one experimental group was not superior to the other in the abilities required for the task. These results are shown in Row 1 of Table 1.

### Overt Levels of Aspiration

Row 2 of Table 1 shows the mean of the scores put before the subjects and said to be originating in *O*. When these scores are com-

pared with the mean overt levels of aspiration (Row 3) it is plain that the overt levels are considerably lower than the scores originating in *O* in both the coercive and referent conditions. The overt levels of aspiration are very similar in the coercive and referent conditions and both are considerably higher than the overt aspirations in the control condition. (Separate inspection of these results trial-by-trial, not shown in the table, revealed that the similarity in overt aspirations between coercive and referent conditions occurred in the first trial and consistently thereafter.) Subjects in the control condition set their overt aspirations only slightly above their obtained performance scores, as is typically found in studies of aspiration setting (Row 5 of Table 1). Coercive and referent subjects, however, placed their overt aspirations much higher than their scores and significantly higher than did the control subjects, suggesting that the social pressures present in the coercive and referent conditions (and absent in the control condition) influenced subjects in these conditions to set higher overt aspirations.

### Self-Evaluation

We have assumed that a person will set his internal level of aspiration in accord with the level originating in *O*, more often in a referent relationship than in a coercive one, and that he will evaluate the quality of his performance in terms of his internal levels of aspiration.

The first and subsequent hypotheses should be assumed to have this phrase preceding them: given a situation in which *Ps* perform below levels originating in *O*:

*Hypothesis 1.* *Ps* in a referent relationship with *O* make a lower evaluation of scores than *Ps* in a coercive relationship with *O*.

In Row 6 of Table 1 it can be seen that subjects who had a referent relationship with *O* were significantly less satisfied with their scores than subjects who had a coercive relationship (and subjects in the control condition). It seems likely therefore that subjects in the coercive relationship and control condition were evaluating their scores on the basis of lower internal levels of aspiration. This last assertion is supported, moreover, in the close similarity between the evaluations of scores made by subjects in the coercive condition and by those in the control condition, indicating that the coerced subjects evaluated their scores on the basis of internal levels of aspiration which were closer to their scores, just as control subjects would be expected to do.

It is clear, moreover, that subjects in the coercive condition did not evaluate their scores by referring to their overt levels of aspiration, since if they had done so the evaluations of their scores would have been considerably lower, given the large discrepancy between their overt levels of aspiration and their scores, shown in Row 5 of Table 1.

### Performance on Test

In the third assumption it was stated that a discrepancy between obtained score and

internal level of aspiration generates forces to decrease this discrepancy. We have seen, in the discussion of Hypothesis 1, that subjects in a referent situation seemed to have set higher internal levels of aspiration, while those under coercion appeared to have set lower levels. Consider, then, the performance scores.

*Hypothesis 2.* *Ps* in a referent situation perform at a higher level than *Ps* in a coercive situation.

The results shown in Row 4 of Table 1 indicate that this hypothesis is supported. Referent subjects scored significantly better than coercive subjects and better than those in the control condition; coercive and control subjects were not significantly different. (On the first trial, a separate analysis revealed, subjects under coercion scored significantly lower than the subjects under referent power and this difference persisted throughout all trials. Indeed, the trial-by-trial scores of subjects in the coercive condition were almost exactly the same as those of subjects in the control condition.)

In essence, the coercive relationship appeared to create weaker motivation to do well. We anticipate, then, that participants in the coercive condition will have felt less motivated on the task.

*Hypothesis 3.* *Ps* in a coercive situation reveal weaker motivation to reach a high level of performance than do *Ps* in a referent situation.

The results in Table 2 provide corroboration for this hypothesis. Subjects in the coercive situation expressed considerably less motivation than subjects in the referent or

TABLE 2  
SUBJECTS STATED MOTIVATION TO PERFORM WELL

Query	Control	Coercive	Referent	<i>t</i> of differences		
	<i>M</i> (A)	<i>M</i> (B)	<i>M</i> (C)	A-B	A-C	B-C
Importance of doing well <sup>a</sup>	5.4	4.4	5.1	2.84**	<i>ns</i>	1.88*
Effort expended to do well <sup>b</sup>	5.8	5.1	6.0	2.50**	<i>ns</i>	3.21***

<sup>a</sup> Query: How important is it to you to do well on this test of small muscle control? (Very little—Very much)  
<sup>b</sup> Query: How hard did you try on the test game-board considering all of your trials here today? (Not at all—Tried my hardest)

\**p* < .05.  
 \*\**p* < .01.  
 \*\*\**p* < .001.



TABLE 3  
ATTITUDES TOWARD SOCIAL PRESSURES ORIGINATING IN O

Query	Control	Coercive	Referent	t of differences		
	M (A)	M (B)	M (C)	A-B	A-C	B-C
Pressure for higher score <sup>a</sup>	3.2	4.2	4.7	2.70*	4.05***	1.35
Others' rating of score <sup>b</sup>	3.6	2.9	3.0	2.50**	2.14*	ns
Lowest score set by others <sup>c</sup>	24.8	29.5	31.5	2.76**	3.92***	1.17
Importance of meeting others' expectations <sup>d</sup>	4.8	4.4	5.6	1.29	2.58**	3.80***
Interest in others' evaluations <sup>e</sup>	4.9	3.8	5.2	2.75**	ns	3.50***

<sup>a</sup> Query: To what extent did you feel there was pressure on you to try for a higher score? (Very little - Very much)

<sup>b</sup> Query: What would probably be the Committee's judgment of your small muscle control? (Very low - Very high)

<sup>c</sup> Query: What do you think the Committee members would set as the lowest score a high school student could get for each trial on the average, and still feel satisfied?

<sup>d</sup> Query: How much did you feel it was important to do as well as the Committee might expect of you? (Very little - Very much)

<sup>e</sup> Query: How much do you care what the Committee thinks about how well you did on the test game board? (Very little - Very much)

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

control conditions. Subjects in referent and control conditions were quite similar in their expressed motivations.

#### *Ps Attitudes toward O's Social Pressures*

The three experimental conditions placed subjects in different positions with respect to actual or perceived pressures from O. We have seen that the internal aspirations of subjects in a referent relationship were apparently more influenced by social pressures originating in O than were the internal aspirations of subjects in a coercive one. Control subjects had no social pressures acting upon them from O and presumably made their decisions about aspirations on the basis of their experiences during the test. How do attitudes toward pressures of O compare in the three conditions? In general, we expect to find that control subjects are neutral toward O's potential pressures since O has exerted no influences upon them and created no conflicts for them, coercive subjects tend to be negative towards O's pressures, and referent subjects tend to be positive toward O's pressures.

First, however, it is useful to examine the degree of pressure subjects felt in these three conditions. Data relevant to this matter are revealed in the upper three rows of Table 3.

In Row 1 it is evident that both coercives and referents felt greater pressure from O to

do well than did subjects in the control condition. The results from two further queries concerning the subjects' perceptions of the committee's judgment of the subjects small muscle control and the lowest score the committee feels a high school student could get for each trial, on the average, and still feel satisfied, can be seen to be essentially supportive of the results just noted for felt strength of pressures. Subjects in the control condition perceived O as making weaker demands of them than did subjects in either of the other two conditions. It is of incidental interest that subjects in both the referent and coercive conditions (all of whom performed below the level originating in O) perceived that O would disapprove of the subjects' performances.

The results in the lower two rows of Table 3 are pertinent to the next hypothesis.

*Hypothesis 4.* Ps in a coercive situation express less acceptance of O's pressures than do Ps in a referent situation.

Subjects in the referent condition state considerably more intention to do as well as O might expect of them and express more concern over the opinion O might have of them, than do subjects in the coercive condition. Control subjects are between these two extremes. Note that subjects in the coercive condition are strikingly low in their concern about the opinion O may have of them.

TABLE 4  
PERCEIVED ATTRACTIVENESS OF OTHER MEMBERS AND GROUP

Query	Control	Coercive	Referent	<i>t</i> of differences		
	<i>M</i> (A)	<i>M</i> (B)	<i>M</i> (C)	A-B	A-C	B-C
Desire to remain on team <sup>a</sup>	5.8	4.3	5.2	3.26***	1.30	1.96*
Liking of others as persons <sup>b</sup>	5.2	4.2	5.5	2.70**	ns	3.51***
Desire to be like others <sup>c</sup>	5.0	4.4	5.7	2.50**	2.91**	5.42***

<sup>a</sup> Query: How much do you want to be a member of the fifteen-man team for which you have been chosen? (Not at all—Very much)

<sup>b</sup> Query: Make your most careful guess about how you would probably feel toward the members of the student committee if you were to meet them (Dislike quite a bit—Like very much)

<sup>c</sup> Query: Do you think the members of the student committee are probably people you would want to be like? Make your best guess.

\*  $p < .05$ .  
\*\*  $p < .01$ .  
\*\*\*  $p < .001$ .

### Attractiveness of Relations with O

*Hypothesis 5.* Ps in a coercive situation perceive O as less attractive than do Ps in a referent situation.

Questions were asked concerning how much subjects desire to remain as members of the team, how much they guess they might like O in future contacts, and how much they desire to be like O. The results from these queries are summarized in Table 4. They reveal that subjects in the coercive condition are less attracted to O than are subjects in either the referent or control conditions. It seems reasonable that participants in the control situation should be attracted to O since, as earlier noted, they are reasonably satisfied with their scores and presume that O approves of their scores. The strong attraction for O presented by referent subjects, and their expressed desire to be similar to O, are indications of their readiness to adopt the standards of performance originating in O, as proposed by French and Raven (1959) and Cartwright (1959). The rejecting reactions revealed by subjects in the coercive situation indicate, as French and Raven and Cartwright have also suggested, that a P who is coerced seeks to avoid a relationship in which coercive power can be used upon him, especially where it is probable that punishment may be forthcoming for his behavior.

### Validity of Score and Centrality of Ability

We have examined results which suggest that subjects in the control and coercive con-

ditions were able to decrease a discrepancy between their levels of aspiration and their scores, while subjects in the referent condition seemed less able to decrease such a discrepancy. We anticipated that the continued forces on referent subjects to decrease this discrepancy generate tendencies to perceive the test as invalid or to perceive the ability involved in the test as less central in the structure of the self.

*Hypothesis 6.* Ps in a referent situation discredit the validity of the test more than those in a coercive situation.

*Hypothesis 7.* Ps in a referent situation discredit the centrality of the tested ability more than those in a coercive situation.

The results to be seen in Table 5 do not support either of these hypotheses. Subjects in the referent and coercive situations are very similar in the degree they discredit the validity of the test and the centrality of the ability. The control subjects, in contrast, are considerably more positive, especially with respect to how central the ability is.

That referent subjects should hold negative opinions about the test and abilities was expected. But why were coercive subjects equally unfavorable in their feelings?

A reasonable explanation is that even though coercive subjects apparently had little though discrepancy between their internal levels of aspiration and their scores, and thus presumably had weak forces on them to discredit the test and the ability, they still had a glaring discrepancy between their scores and the



TABLE 5  
ATTITUDES TOWARD VALIDITY OF SCORE AND CENTRALITY OF ABILITY

Query	Control	Coercive	Referent	t of differences		
	M (A)	M (B)	M (C)	A-B	A-C	B-C
Validity of score <sup>a</sup>	4.5	3.6	4.0	2.43*	1.35	1.08
Centrality of ability <sup>b</sup>	6.2	5.4	5.4	2.58**	2.58**	ns

<sup>a</sup> Query: Do you think your score on this test truly shows how much small muscle control you have? (Definitely not—Definitely yes)

<sup>b</sup> Query: In your opinion, how important or unimportant is it for you to be good in small muscle control?

\*  $p < .05$ .

\*\*  $p < .01$ .

standards put before them by O and an apparent discrepancy between their internal levels of aspiration and the standards originating in O. These latter two discrepancies perhaps caused the coercive subjects to perceive a strong likelihood of disapproval from O. Subjects under coercion, therefore, in seeking to discredit the test and the centrality of the ability were attempting to weaken the impact of the probable disapproval and rejection from the team by trying to convince themselves that the testing procedure was invalid, and the tested ability less important than the experimenter had earlier asserted.

#### Emotional Tension

Tension was anticipated in both the referent and coercive subjects because of forces on them to reduce the discrepancy between performance and internal level of aspiration, opposed by felt pressures to raise the level of aspiration. Control subjects, however, are assumed to have been under little emotional tension.

*Hypothesis 8.* Emotional tension will be greater in Ps in both the referent and coercive conditions than in the control condition.

The results in Table 6 corroborate this hypothesis. Subjects in both coercive and referent situations felt more worried during the test and alleged that they were more often too tense to shoot well than did subjects in the control condition. With respect to "worry," the coercive and referent subjects were significantly different from one another. The amount of emotional tension expressed by referent subjects was greater than that mentioned by control subjects, with coercive

subjects falling between these two extremes. The results suggest that referent subjects were under more tension than were coercive subjects.

#### DISCUSSION

The second assumption, in the opening paragraphs, was that referent social power would be more effective in changing private cognitions, and thus in determining the location of the internal levels of aspiration, than coercive power would be. The results indicate that this assumption appears to have been warranted. It is based, however, on a further (working) assumption that referent and coercive social power were approximately equal in strength. There is no way in the present design accurately to determine whether this equality existed between the two types of power. The fact that subjects were fairly alike when describing the pressures they felt from O in the referent and coercive conditions (see Table 3) suggests that the pressures in the two conditions may have been approximately equal. It remains for future study to determine if weak referent power, for example, compared to very strong coercive power, as well as comparisons of other combinations of strength in the two types of power, would obtain results generally supportive of those reported here.

Our expectation was that withdrawal from the situation was to be anticipated of subjects in the coercive condition because of the threats of punishment: rejection from the group, loss of participation in the state tournament, as well as loss of self-esteem. It is possible, however, that we confounded matters

by predicting withdrawal in a coercive situation where the potential punishment implied rejection of the subject and by predicting non-withdrawal in a referent situation where the subject was assured of his membership on the team. One wonders if results similar to those presently described would be obtained if coercion had been operationalized by potential fines or by some other deprivation than threatened rejection from the group. Also, what would occur if coercive social pressures had been made operational by a large reward for appropriate behavior? Clearly different ways of making the independent variables operational need careful study before we can be sure how far the present results may be generalized as consequences of referent and coercive power.

## SUMMARY

Individual subjects were observed in a situation similar to the one used by Rotter (1942) for studying the level of aspiration, while they were to perceive themselves as members of a team which was not physically present. In two experimental conditions (referent and coercive) the subjects were provided standards of individual performance, for a series of trials, said to originate in other members of their team. Events were so arranged that subjects, over the series, did not attain the performance levels put before them by the others. In a third condition (control) subjects were not provided standards of performance from others. Assumptions were made that: (a) scores are evaluated in terms of internal levels of aspiration; (b) referent social pressures more strongly influence cog-

nitions (e.g., internal levels of aspiration) than do coercive social pressures; and (c) a discrepancy between an internal level of aspiration and an obtained score generates forces to reduce this discrepancy.

Several hypotheses concerning self-evaluation and performance were corroborated, where all subjects scored below the levels proposed for them by others.

*Hypothesis 1.* *Ps* in a referent relationship with *O* make a lower evaluation of their scores than do *Ps* in a coercive relationship with *O*.

*Hypothesis 2.* *Ps* in a referent situation obtain higher scores than do *Ps* in a coercive situation.

*Hypothesis 3.* *Ps* in a coercive situation reveal weaker motivation to reach a high score than do *Ps* in a referent situation.

Several hypotheses, comparing the reactions of failing subjects who are in coercive and referent relation with others, receive support.

*Hypothesis 4.* *Ps* in a coercive situation perceive *O* as less attractive than do *Ps* in a referent situation.

*Hypothesis 5.* *Ps* in a coercive situation state less acceptance of *O*'s social pressures than do *Ps* in a referent situation.

*Hypothesis 6.* Emotional tension is greater in *Ps* either in referent or coercive situations than in *Ps* under no social pressures in the control condition.

Two other hypotheses were not confirmed.

*Hypothesis 7.* *Ps* in a referent situation discredit the validity of the test more than do *Ps* in a coercive situation.

*Hypothesis 8.* *Ps* in a referent situation discredit the centrality of the tested ability more than do *Ps* in a coercive situation.

TABLE 6  
EMOTIONAL TENSION INDICATED BY SUBJECTS

Query				t of differences		
	Control	Coercive	Referent	A-B	A-C	B-C
	<i>M</i> (A)	<i>M</i> (B)	<i>M</i> (C)			
Worried during session <sup>a</sup>	2.8	3.7	4.1	2.90**	5.16***	2.26*
Tense during session <sup>b</sup>	3.0	3.8	4.4	1.90*	3.33***	1.43

<sup>a</sup> Query: Did you notice yourself becoming worried at any time during this test-taking session? (Not at all—Quite a bit)

<sup>b</sup> Query: When making your shots, how often did you feel you were too tense to give your best performance? (Hardly ever—

Very often)

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .



## REFERENCES

- CARTWRIGHT, D. A field theoretical conception of power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Institute for Social Research, 1959. Pp. 183-220.
- CHAPMAN, D., & VOLKMANN, J. A social determinant of the level of aspiration. *J. abnorm. soc. Psychol.*, 1939, **34**, 225-238.
- FESTINGER, L. Wish, expectation, and standards as factors influencing level of aspiration. *J. abnorm. soc. Psychol.*, 1942, **37**, 184-200.
- FRENCH, J. R. P., JR., & RAVEN, B. The bases of social power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor, Mich.: Institute for Social Research, 1959. Pp. 150-167.
- LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald, 1944. Pp. 333-378.
- LINDQUIST, E. *Design and analysis of experiments in psychology and education*. Boston, Mass.: Houghton Mifflin, 1953.
- RAVEN, B. Social influences on opinion and the communication of related content. *J. abnorm. soc. Psychol.*, 1959, **58**, 119-128.
- ROTTER, J. Level of aspiration as a method of studying personality. *J. exp. Psychol.*, 1942, **31**, 410-422.
- STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55-63.
- STOTLAND, E., & ZANDER, A. Effects of public and private failure on self-evaluation. *J. abnorm. soc. Psychol.*, 1958, **56**, 223-229.
- ZANDER, A., COHEN, A., & STOTLAND, E. *Role relations in the mental health professions*. Ann Arbor, Mich.: Institute for Social Research, 1957.
- ZANDER, A., NATSOULAS, T., & THOMAS, E. Personal goals and the group's goals for the member. *Hum. Relat.*, 1960, **13**, 333-344.

(Received January 20, 1961)

## CRITIQUE AND NOTES

### PERSONALITY CORRELATES OF PICTURE PREFERENCES<sup>1</sup>

GEORGE CERBUS<sup>2</sup> AND ROBERT C. NICHOLS

*Purdue University*

#### METHOD

The use of esthetic preferences as measures of personality would be expected to combine the subtlety of projective tests with the adaptability to statistical treatment of the personality questionnaire. Barron (1953), Christensen (1960), Knapp and Green (1959), and Welsh (1959) have demonstrated significant relationships between personality measures and preferences for abstract figures and paintings. Investigations by Eysenck (1941a, 1941b), Barry (1957), Weingarten (1958), and Shooster (1957) have indicated that preferences for representational pictorial art have significant personality correlates, especially with interests.

In order to explore more thoroughly the potentialities of this technique for personality assessment, three scales were constructed to measure preferences for qualities of pictorial art, and the relationships of these preferences to personality measures were investigated. The first variable, achromatic-chromatic qualities of clinical and experimental consideration of perceptions involving color on the Rorschach, as well as by the finding of an achromatic factor by Knapp and Green (1959) as the first of five obtained from factor analysis of preferences for abstract paintings. The second variable, abstract versus representational or concrete qualities, was suggested by Eysenck's (1941a) early description of the second factor obtained by factor analysis of ratings of a variety of pictures. This variable would also be quite similar to the complexity-simplicity variable studied by Barron (1953). The third variable, human subjects versus depiction of nonhuman objects, was chosen to investigate the use of picture content in personality assessment, suggested by the promising results of the pictorial forms of interest tests developed by Weingarten (1958) and Shooster (1957).

Ninety-nine slides of a wide variety of artistic productions, most of which were not widely known, were chosen from a file of approximately 10,000 slides in the Art Design Department of a university. Each picture was individually judged by the experimenter on the three variables chosen for investigation. In an attempt to minimize the general preference factor found by Welsh (1949) and Eysenck (1941a) to be the major contributor to the variance in artistic preferences, each picture was then rated by a group of 25 student judges on a five-point scale of general esthetic appeal. Pairs of pictures of equivalent esthetic preference were then chosen with opposing characteristics on one variable and common characteristics on the remaining two variables. In all subsequent testing, the pictures were projected on a screen in pairs and the subjects were instructed to indicate which of the two pictures they preferred.

Preliminary scales of the three variables, each of 58 items (pairs of pictures), were selected and administered to a group of 106 neuropsychiatric patients in a Veterans Administration hospital. An item analysis was carried out on each scale, following a procedure described by Webster (1956), in which those items with low correlations with the total score were discarded. Reliabilities of the improved scales were determined on a sample of 118 college students enrolled in introductory psychology, who also served as a population for a second item analysis of each scale. The final scales, consisting of items surviving both item analyses, should have internal consistency in diverse groups.

For indications of personality correlates of these three variables, scores of the improved scales were correlated with scores on the California Psychological Inventory (CPI) in the college sample, with scores on the Minnesota Multiphasic Personality Inventory on the Minnesota population, and with (MMPI) in the hospital population, and with therapists' ratings on the Multidimensional Scale for Rating Psychiatric Patients (MSRPP) (Lorr, 1953) in an additional sample of 51 neuropsychiatric patients. The 51 therapy cases had all been seen for at least four group therapy sessions in a Veterans Administration hospital.

#### RESULTS

##### *Achromatic-Chromatic Scale*

After elimination of items as a result of the item analyses, 27 items were used as the final form of this scale. The shorter Kuder-Richard-

<sup>1</sup> This paper is based in part on a doctoral dissertation submitted by the senior author to the faculty of Purdue University in January 1961.

<sup>2</sup> Now at the Marshall County Mental Health Center, Marshalltown, Iowa.



TABLE 1

CORRELATIONS OF THE PICTURE PREFERENCE SCALES  
WITH SCALES OF THE MMPI

(N = 98 neuropsychiatric patients)

MMPI scales	Achromatic-Chromatic scale <sup>a</sup>	Concrete-Abstract scale <sup>b</sup>	Objects-People scale <sup>c</sup>
<i>L</i>	.02	.18	.04
<i>F</i>	.04	-.28**	.17
<i>K</i>	.03	.12	-.05
<i>Hs</i>	.34**	.06	.15
<i>D</i>	.10	-.10	.20*
<i>Hy</i>	.20*	.05	.17
<i>Pd</i>	.19	-.08	-.03
<i>Mf</i>	-.18	-.23*	-.23*
<i>Pa</i>	.01	-.16	.11
<i>Pt</i>	.06	-.13	.11
<i>Sc</i>	.01	-.24*	.15
<i>Ma</i>	.09	-.19	.02

<sup>a</sup> *M* = 13.5, *SD* = 5.3.<sup>b</sup> *M* = 32.1, *SD* = 7.8.<sup>c</sup> *M* = 15.3, *SD* = 5.5.

\* Significant at the .05 level.

\*\* Significant at the .01 level.

son Formula 21 yielded a cross-validated coefficient of .71 for a 31-item form of this scale in the sample of 118 college students. Subsequently, four items were discarded as a result of the second item analysis. The scoring direction was such that numerically high scores indicated preference for achromatic pictures.

Correlations with the other picture preference scales developed in this study were low and generally not significant. In the combined sample of 224 subjects, it was not significantly related to the Concrete-Abstract scale (correlation of .11), nor to the Objects-People scale (-.04). A significant deviation from linearity was found in the regression of Achromatic-Chromatic and Objects-People scores, and the correlation ratio of .35 was significant at the .01 level. Distribution of scores on the two scales was relatively normal, but subjects preferring colored pictures tended to score either high or low on the Objects-People scale.

Significant correlations with other personality measures employed here were limited to those with the Hypochondriasis (.34) and Hysteria (.20) scales of the MMPI (see Table 1). There were no significant correlations with CPI scales nor with therapists' ratings on the MSRPP.

#### Concrete-Abstract Scale

The item analyses resulted in use of 42 items in the final form of this scale, which was found to have highly satisfactory reliability. Using the sample of college students and a preliminary 45-item form of the scale derived on the

hospital sample, the coefficient of reliability was .89 as estimated by Kuder-Richardson Formula 21. The correlation with the Objects-People scale was low but significant (.14 with 224 subjects, significant at the .05 level). Numerically high scores on this scale indicate preference for concrete pictures.

Significant correlations with verbal personality measures were limited to correlations with the *F* (-.28), *Mf* (-.23), and *Sc* (-.24) scales of the MMPI (see Table 2). Only one correlation with psychotherapists' ratings attained significance, a point-biserial correlation of -.34 with ratings on presence of "impulses to commit a criminal or hostile act" (see Table 3, Item 16).

The significant relationship between preference for abstract paintings and many unusual responses to verbal items of a questionnaire (the *F* scale of the MMPI) suggested that the preference for abstract paintings is not prevalent. This is corroborated by the finding of a mean Concrete-Abstract score (28.8 with 224 subjects) significantly higher, at well beyond the .01 level, than the assumed chance score of 21 in a 42-item scale. This dislike for the abstract painting was more marked in the patient population, for the mean among patients (32.8) was significantly higher at the .01 level than the

TABLE 2

CORRELATIONS OF THE PICTURE PREFERENCE SCALES  
WITH SCALES OF THE CPI

(N = 117 college students)

CPI scales	Achromatic-Chromatic scale <sup>a</sup>	Concrete-Abstract scale <sup>b</sup>	Objects-People scale <sup>c</sup>
Do	.04	.11	-.24**
Cs	.05	-.13	-.22*
Sy	.09	-.01	-.16
Sp	-.02	.03	-.22*
Sa	.08	-.08	-.23*
Wb	-.09	.11	-.02
Re	-.04	-.09	-.07
So	-.12	-.02	.21*
Sc	-.03	.04	.17
To	-.14	.01	-.04
Gi	.08	-.01	-.06
Cm	-.14	-.12	-.02
Ac	-.02	.04	-.10
Ai	-.07	-.01	-.08
Ie	-.01	.07	-.10
Py	-.03	.08	-.10
Fx	.05	-.07	-.08
Fe	-.03	-.13	.10

<sup>a</sup> *M* = 12.0, *SD* = 4.9.<sup>b</sup> *M* = 25.6, *SD* = 9.4.<sup>c</sup> *M* = 15.8, *SD* = 5.8.

\* Significant at the .05 level.

\*\* Significant at the .01 level.

TABLE 3

CORRELATIONS OF THE PICTURE PREFERENCE SCALES WITH SCALES AND FACTORS OF THE MSRPP  
( $N = 51$  neuropsychiatric patients)

Scale number	Characteristic given high rating	Achromatic-Chromatic scale	Concrete-Abstract scale	Objects-People scale
5.	Tense	-.02	.00	.04
6.	Conviction of sinfulness <sup>a</sup>	.02	.04	.33*
8.	Overtalkative	.05	-.07	-.17
10.	Believes he is influenced <sup>a</sup>	-.16	-.18	-.28*
11.	Disoriented for people <sup>a</sup>	.15	.20	.11
16.	Disturbed by hostile impulses <sup>a</sup>	-.12	-.34*	-.23
17.	Paralyzing anxiety	-.04	.04	.16
19.	Frequent use of complaints <sup>a</sup>	.03	-.08	-.04
20.	Shouts or yells	.06	-.08	-.27
21.	Extreme body concern <sup>a</sup>	-.04	-.04	.00
22.	Hallucinatory voices <sup>a</sup>	-.05	.15	.12
23.	Frequent mood swings	-.07	-.07	-.07
25.	Conceited	-.02	-.06	-.46**
27.	Hallucinatory visions <sup>a</sup>	-.11	.12	.17
30.	Grandiose <sup>a</sup>	.05	.00	-.32*
31.	Uninhibited	.08	.09	-.12
33.	Conviction of persecution	-.16	.07	-.08
34.	Elated	.03	-.18	-.31*
35.	Suicidal preoccupation <sup>a</sup>	-.02	.06	.23
36.	Overresponsive	.01	-.03	-.16
37.	Ideas of reference	-.14	-.08	-.06
38.	Impossible beliefs	-.05	-.06	-.17
40.	Anticipates disasters	-.10	-.14	.08
53.	Neat and clean <sup>a</sup>	.03	.02	-.14
56.	Always talking to others	.08	-.18	-.29*
57.	Many interests	.03	-.17	-.38**
58.	Sociable	-.02	-.02	-.25
Factors				
A.	Retarded depression versus manic excitement	.04	-.06	-.24
C.	Paranoid projection <sup>a</sup>	-.20	-.13	-.16
E.	Melancholy agitation	-.08	.01	.08
F.	Perceptual distortion <sup>a</sup>	-.06	-.19	-.14
I.	Withdrawal	.04	-.13	-.37**
J.	Self-depreciation versus expansiveness	-.02	-.07	-.42**

Note.—Only the items shown were rated. These items were chosen from the total scale as being most likely to be related to the picture preference scales after studying the correlations with the inventory tests.

<sup>a</sup> Because of skewed distributions, the ratings on these items were dichotomized and the resultant point-biserial correlations are listed.

\* Significant at the .05 level.

\*\* Significant at the .01 level.

mean among college students (25.7). Additionally, there was a significant difference among patients, for the mean among those diagnosed neurotic (37.0) was significantly higher than the mean among those diagnosed psychotic (30.5).

#### Objects-People Scale

Item analyses resulted in use of 33 items in the final form of this scale. The Kuder-Richardson Formula 21 yielded an estimate of .74 for the reliability coefficient of a preliminary 39-item form of this scale in the cross-validating sample. Numerically high scores on this scale indicate preference for pictures depicting objects. Seven correlations with verbal personality

scales were significant. Four of these were with CPI scales placed by Gough (1957) in a group called "Measures of Poise, Ascendancy, and Self-Assurance." These are the scales of Dominance (-.24), Capacity for Status (-.22), Social Presence (-.22), and Self-Acceptance (-.23). Additionally, the correlations with the Socialization scale of the CPI (.21) and the *D* (.20) and *Mf* (-.23) scales of the MMPI attained significance (see Tables 1 and 2).

Scores on the Objects-People scale correlated significantly with nine measures on the MSRPP (see Table 3). The preference for pictures of objects was significantly related to factor scores of withdrawal (-.37, with low factor scores indicating withdrawal) and self-depreciation



versus grandiose expansiveness ( $-.42$ , with low scores indicating self-depreciation), with five individual scales included in those factors, and with additional scales of belief in being influenced or controlled by others (Item 10 of the MSRPP,  $-.28$ ) and of depression (Item 34,  $-.31$ ).

#### DISCUSSION

In this study, in which each of the three experimental scales is correlated with 63 other variables, a finding of more than three significant correlations should be interpreted as a suggestive rather than a definitive finding.

The finding of two significant correlations with preference for achromatic over chromatic pictures should be attributed to chance. This was surprising to the present investigators because of the great significance attached to use of color in the Rorschach. It is possible, however, that use of color in Rorschach perceptions is a different phenomenon from the influence of color in esthetic preferences. The fact that more difficulty was experienced in constructing a reliable scale for achromatic preference than for the other two variables suggests that color may not be a very potent factor in determining esthetic preferences.

The finding of four significant correlations with preference for concrete over abstract pictures could be attributed to chance. However, these correlations along with the significant group differences noted above are consistent with the interpretation that those who prefer the concrete to the abstract pictures are the more conforming and defensive. This is what one would expect from Barron's (1953) study of complexity-simplicity.

The Objects-People scale had significant relationships with 16 of the other variables. These correlations suggest that the person who prefers pictures containing people to those without people is more outgoing and friendly, and tends to be more expansive, self-confident, and happy than the person who prefers pictures of objects.

#### SUMMARY

Three picture preference scales based on judgments of achromatic-chromatic and concrete-

abstract qualities and content of pictures were constructed by pairing pictures with opposing characteristics. Scales of 27, 42, and 33 items were found to have satisfactory reliabilities (.71 to .89), with generally insignificant linear inter-correlations.

Correlations of each of the picture preference scales with 63 verbal personality measures and personality ratings yielded results differing little from chance with the two scales based on structural qualities, contrasting sharply with the 16 significant correlations found with the scale based on picture content, i.e., pictures depicting objects versus pictures of people.

#### REFERENCES

- BARRON, F. Complexity-simplicity as a personality dimension. *J. abnorm. soc. Psychol.*, 1953, **48**, 163-173.
- BARRY, H. Relationships between child training and the pictorial arts. *J. abnorm. soc. Psychol.*, 1957, **54**, 380-383.
- CHRISTENSEN, C. M. Dimensions and personality correlates of abstract design preferences. *Amer. Psychologist*, 1960, **15**, 453. (Abstract)
- EYSENCK, H. J. The general factor in aesthetic judgements. *Brit. J. Psychol.*, 1941, **31**, 94-102. (a)
- EYSENCK, H. J. "Type"-factors in aesthetic judgements. *Brit. J. Psychol.*, 1941, **31**, 262-270. (b)
- GOUGH, H. G. *California Personality Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
- KNAPP, R. H., & GREEN, S. M. Personality correlates of preferences for abstract paintings. *Amer. Psychologist*, 1959, **14**, 392. (Abstract)
- LORR, M. Multidimensional Scale for Rating Psychiatric Patients. *Veterans Admin. tech. Bull.*, November 1953, No. 10-507.
- SHOOSTER, C. *Pictorial study of values*. Chicago, Ill.: Psychometric Affiliates, 1957.
- WEBSTER, H. Transformed statistics for use in test construction. *Psychol. Bull.*, 1956, **53**, 488-492.
- WEINGARTEN, K. P. *Picture interest inventory*. Los Angeles, Calif.: California Test Bureau, 1958.
- WELSH, G. S. A projective figure-preference test for diagnosis of psychopathology. Unpublished doctoral dissertation, University of Minnesota, 1949.
- WELSH, G. S. *Preliminary manual for the Welsh figure preference test*. Palo Alto, Calif.: Consulting Psychologists Press, 1959.

(Received January 3, 1961)

# EFFECT OF DISTRACTING WORDS ON A SIZE ESTIMATION TASK

COLEMAN PAUL<sup>1</sup>

*Center for the Study of Cognitive Processes, Wayne State University*

The effects of motivational variables upon perception have received considerable attention in recent experimentation. One typical approach to the problem of how needs affect perception is illustrated by the size estimation study of Bruner and Postman (1948) in which the subjects had to adjust a variable disc of light to match a standard disc. Printed on the standard disc was a "value symbol" such as a dollar sign, a swastika, etc. This study indicated that the disc size tended to be overestimated when it bore a positive value symbol. Klein, Schlesinger, and Meister (1951) employed more stringent controls and did not corroborate this finding. A second approach to the same problem was taken by McGinnies (1949), who measured the subjects' recognition thresholds for different classes of words (i.e., tabu, neutral, and reward). He reported a higher threshold for tabu than for neutral words and elaborated a perceptual defense concept. Modifications of McGinnies' study have adduced evidence that response suppression (Whittaker, Gilchrist, & Fischer, 1952) and frequency of occurrence of the words (Solomon & Postman, 1952) are variables with greater explanatory value than the perceptual defense hypothesis.

The present study controlled the response suppression, frequency of occurrence, and configuration of stimulus variables. Using the method of absolute judgment, subjects had to estimate the lengths of various lines and to ignore words which were printed adjacent to the lines. Words considered to have reward, tabu, or neutral connotations were employed and were hypothesized to exercise differential distraction effects resulting in differential errors in size estimation.

## METHOD

**Subjects.** Ten students, male and female, within the 19-24 year age range served as subjects.

**Apparatus.** Lines 1.5, 2, 3, 4, and 5 inches long and 1/16 inch wide were each paired once with words having associations that were tabu, neutral, or reward and each length appeared once without any word to control for effects of words per se. Each word was printed in 1-inch high letters and was .5 inch from the line, randomly to the right or left of the line.

The stimulus cards were presented for a duration of .1 second by means of a two-field Dodge-type tachistoscope. This exposure time allowed easy recognition of the stimulus card.

**Procedure.** The 50 stimulus cards were presented in a fixed random order to each subject. The subject was instructed to attend to the exposure field when given a "ready" signal and to report the length of line presented. After the verbal report, the subject was to make a freehand drawing of the line and the drawing was then removed. The subjects were told neither the lengths of lines nor the number of different lines being used.

## RESULTS

The discrepancy between the subjects' reports and actual line lengths was calculated and a constant of 5 was added to make each measure positive. The mean for the three words of each condition (tabu, neutral, and reward) and the control were taken as criterion measures. The verbal reports and the drawings were treated separately as "modes."

An analysis of variance (Word Conditions  $\times$  Lengths  $\times$  Modes  $\times$  Subjects) was performed (Table 1). The main effect for conditions was significant. Examination of the data, however, indicated that the significant difference was between the control and word conditions, not between the different word conditions (Table 2).

TABLE 1  
ANALYSIS OF VARIANCE OF ERRORS OF ESTIMATION

Source	df	MS	F
Modes (M)	1	1.03	0.83
Conditions (C)	3	1.16	4.00**
Lengths (L)	4	8.52	5.95**
Subjects (S)	9	17.35	78.86***
M $\times$ C	3	0.09	1.50
M $\times$ S	9	1.24	8.27***
M $\times$ L	4	0.42	2.80*
C $\times$ S	27	0.28	1.27
C $\times$ L	12	0.35	1.59
L $\times$ S	36	1.43	6.50***
L $\times$ M $\times$ S	36	0.15	0.94
L $\times$ C $\times$ S	108	0.22	1.38
C $\times$ M $\times$ L	12	0.09	0.58
C $\times$ M $\times$ S	27	0.06	0.37
L $\times$ M $\times$ C $\times$ S	108	0.16	—

\*  $p < .05$ .

\*\*  $p < .025$ .

\*\*\*  $p < .001$ .

<sup>1</sup> Now at Lafayette Clinic, Detroit, Michigan.



TABLE 2  
MEANS FOR WORD CONDITIONS BY MODES AND FOR TOTAL WORD CONDITIONS

		Tabu	Neutral	Reward	Control	
Verbal	$\bar{X}$	4.5	4.6	4.5	4.3	$T \times N = 0.44$
(Ns = 50)	$\sigma$	0.91	0.96	1.05	0.81	$T \times R = 0.21$
						$T \times C = 1.29$
						$N \times R = 0.20$
						$N \times C = 1.67$
						$R \times C = 1.39$
Motor	$\bar{X}$	4.4	4.4	4.4	4.2	$T \times N = 0.06$
(Ns = 50)	$\sigma$	0.83	0.84	0.95	0.71	$T \times R = 0.28$
						$T \times C = 1.33$
						$N \times R = 0.22$
						$N \times C = 1.22$
						$R \times C = 0.90$
Total	$\bar{X}$	4.5	4.5	4.5	4.3	$T \times N = 0.25$
(Ns = 100)	$\sigma$	0.87	0.90	1.00	0.76	$T \times R = 0.07$
						$T \times C = 1.83^*$
						$N \times R = 0.31$
						$N \times C = 2.04^{**}$
						$R \times C = 1.60$

\*  $p < .10$ .

\*\*  $p < .05$ .

As would be expected, the main effects for subjects and lengths were significant. The Modes  $\times$  Subjects interaction indicated that subjects differed in the accuracy with which they verbalized or drew the length of line. The Modes  $\times$  Lengths interaction indicated that some lengths were reported more accurately by verbal response and some more accurately by motor response.

#### SUMMARY

In an experimental task involving size estimation, the contiguous presentation of meaningful, but task irrelevant, stimuli have little effect. It appears that the structural properties of irrelevant stimuli have greater effects than do "meaning" properties when the latter are task irrelevant.

#### REFERENCES

- BRUNER, J. S., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. abnorm. soc. Psychol.*, 1948, 27, 203-208.
- KLEIN, G. S., SCHLESINGER, H. J., & MEISTER, D. E. The effect of personal values on perception: An experimental critique. *Psychol. Rev.*, 1951, 58, 96-112.
- MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, 56, 244-251.
- SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, 43, 195-201.
- WHITTAKER, E. M., GILCHRIST, J. C., & FISCHER, J. W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, 47, 732-733.

(Received January 3, 1961)

## SPONTANEOUS FANTASY AS A RESOURCE OF HIGH GRADE RETARDATES FOR COPING WITH A FAILURE-STRESS FRUSTRATION<sup>1</sup>

GEORGE H. SAXTON<sup>2</sup>

*Edward R. Johnstone Training and Research Center, Bordentown, New Jersey*

Although fantasy, as elicited by various projective techniques, has been studied rather extensively, spontaneous fantasy, or daydreaming, has received little attention. The dearth of scientific research in this area is unfortunate and surprising in light of the importance placed on daydreaming by clinicians and personality theorists. Furthermore, there may be need for more reservation in making inferences regarding a subject's fantasy life from his performance on projective techniques (Page, 1956, 1957).

Shaffer and Shoben (1956) have demonstrated that daydreaming may constitute a far more frequent phenomenon in a normal population than was previously recognized. Page and Epstein (1953 unpublished) developed a Scale of Fantasy Behavior which assesses various daydreaming habits and themes, and collected further normative data which basically supports Shaffer and Shoben's findings. Page (1956, 1957) compared high fantasy (HF) and low fantasy (LF) groups, as defined by his Fantasy Scale, on TAT and Rorschach protocols. On the Rorschach, the groups were found to differ significantly in total number of responses, and in the *M* production when the determinant was scored very rigorously, i.e., popular *M* percepts were eliminated from consideration. In both cases, the HF group was superior. There was no significant difference between groups on TAT productivity or on ratings of emotional tone, and some inverse relationships were obtained between TAT and daydreaming themes, i.e., low incidence of heterosexual daydreams was correlated with high incidence of heterosexual themes on the TAT protocols.

Many studies have demonstrated that the effects of various types of unpleasant situations

are reflected in fantasy behavior. Sears (1942) observed a decrease in general motility level and a sharp increase in daydreaming and autistic behavior after subjecting subjects to a failure situation. Bellak (1944) demonstrated that the number of aggressive words increased in TAT stories when subjects were criticized for inferior performance. Fantasy aggression has been shown to be positively correlated with overt aggression in some studies (Kagan, 1956; Lesser, 1957), and negatively correlated in others (Feshbach, 1955). In the latter study, Feshbach hypothesized that the expression of aggression in fantasy will serve partially to reduce aggressive drive. His procedure involved insulting a group of college students in a classroom situation, and then introducing a fantasy activity (TAT) and a nonfantasy activity after which measures of aggressive drive were obtained. Results primarily supported the research hypothesis. Feshbach stated that, "As viewed here, fantasy behavior is an adjustment mechanism which can serve to reduce tensions and provide substitute goal satisfactions. It may function as an outlet for socially unacceptable motives and frustrated achievement strivings" (p. 10). Furthermore, regarding the possible differences between spontaneous and elicited fantasy, Feshbach stated that the former may be more effective in reducing aggressive drive.

This study is an attempt to investigate the hypothesis that retardates characterized by high utilization of fantasy as measured by the Fantasy Inventory, will manifest less aggression, and situational anxiety, than retardates characterized by low utilization of fantasy. Beier, Gorlow, and Stacey (1951), in studying the fantasy life of the mental defective, have demonstrated that this population can yield surprisingly rich and productive TAT protocols, and Hackbusch and Klopfer (1946) and Sarason (1945), have also stressed the importance of studying fantasy behavior in these subjects.

### METHOD

*Subjects.* The subjects consisted of 86 institutionalized male and female high grade retardates between the ages of 12 and 22 years. All subjects were

<sup>1</sup> Submitted to the Temple University Graduate Council in partial fulfillment of the requirements for the PhD degree. The research was conducted at the Edward R. Johnstone Training and Research Center where the author was a Research Fellow.

<sup>2</sup> Sincere thanks are extended to Leonard Blackman, Ronald Lipman, Herman Spitz, and Belver Griffith of the Johnstone Center, and to Horace Page, Harvey Austrin, and Harold Reppert of Temple University.



tested by the institutional psychology staff with the Wechsler Intelligence Scale for Children for those subjects under 16 years, and the Wechsler Adult Intelligence Scale for those subjects 16 years and over. Only subjects with a Full Scale IQ of 60 or above were accepted for this study. All IQs were current within one year. Subjects diagnosed as psychotic were rejected.

*Procedure.* Testing was done in two sessions, all tests being individually administered. The first session involved administration of the following:

1. Fantasy Inventory which consisted of 50 items with vocabulary modifications appropriate for retardates, drawn from Page and Epstein's (1953 unpublished) Scale of Fantasy Behavior. (A split-half reliability coefficient of .90 had been obtained with the Page-Epstein scale, and an estimate of validity of .48 by correlating ratings of day-to-day journals of daydreaming behavior with the Fantasy Scale. The Fantasy Inventory for mental retardates was found to have a Spearman-Brown reliability coefficient of .89.)

2. A shortened version of the Children's form of the Manifest Anxiety (CMA) scale developed by Castaneda, McCandless, and Palermo (1956), was utilized to offer some indication of basic anxiety level of the subjects and study its possible relationship with the fantasy variable.

3. Digit Span test (Wechsler-Bellevue Intelligence Scale, I) to obtain a prefrustration estimate of digit span performance which will serve as a baseline for comparison with a postfrustration digit span test measuring intellectual response to situational anxiety.

HF and LF groups were selected by scores on the Fantasy Inventory. Both the HF and LF groups were then broken down randomly into experimental (frustration) and control (nonfrustration) treatment groups for the second session's testing. The control (nonfrustrated) group was introduced to evaluate the efficacy of the experimental introduction of frustration on the subjects.

The second session involved subjecting both experimental and control groups to a task in which success or failure was a highly subjective matter, and could be arbitrarily decided by the examiner. This task consisted of having subjects trace around various geometric patterns, staying neatly between two lines, approximately .25 inch apart. The Mirror Drawing apparatus was utilized with the experimental group, but not with the control group, due to the stressful nature of the task. Control subjects who finished before 5 minutes were asked to connect a series of dots with neat, straight lines, alternating with a red and blue pencil until 5 minutes had elapsed. Both groups were told that they were being timed, and that performance was evaluated for quantity and quality. Subjects were then presented with a reward (a piece of candy or a cigarette, depending on the subject's preference) and told that this reward, as well as two additional rewards, could be earned if they did well; however, if they did poorly even the initial reward would be withdrawn. Subjects were then told to begin the tasks.

The following procedure was used with the experimental group: At the end of 3 minutes work the experimenter said, "You are not doing very well. You must hurry up." After 4 minutes the experimenter said, "Your work is very sloppy and you are too slow. Please try to do better." After 5 minutes, the subject's work was collected and the experimenter said, "Practically everyone did better than this. I will have to take your reward back as you did not earn it." The procedure for the control group was as follows: At the end of 3 minutes the experimenter said, "You are doing well. Keep up the good work." At 4 minutes the subject was presented with an additional reward and was told, "You are doing very good work. You have earned another reward." At 5 minutes the subject's work was collected and he was presented with a third reward and told, "You did very well. You have earned three rewards altogether."

At this point, both experimental and control subjects had a 5-minute wait period, during which time they were instructed to sit quietly in a large cushioned chair in the testing room. During this period overt behavior such as talking, smoking, and chewing, was restricted for experimental subjects. Control subjects were permitted to smoke or eat candy if they desired. The purpose of this period was to allow subjects time to engage in fantasy behavior, if they were so inclined.

The following measures of the dependent variable were then taken:

1. Attitude Questionnaire, designed to assess aggressive negativistic feelings toward: the examiner, research, Johnstone personnel, and impersonal aspects of the Johnstone Center. (This consisted of 25 items, to be answered "Yes" or "No" by the subjects.)

2. Digit Span test (Wechsler-Bellevue Intelligence Scale, II).

3. Picture Technique, consisting of three sketches which varied in their likeness to the experimenter. (Subjects ranked the sketches from "most" to "least" liked.)

The attitude questionnaire and picture technique were administered by other researchers, and not the original (frustrating) experimenter. On the basis of the research hypothesis it was predicted that after the failure-stress frustration the HF group would manifest a more positive attitude, show less of a decrement in digit span performance, and rank the examiner sketch more favorably, than the LF group. The experimental and control treatment groups were compared on all criteria measures to evaluate the effects of the frustration (experimental) treatment.

## RESULTS

In the first session's results, the HF and LF groups (fantasy scores were normally distributed) were found not to differ significantly with respect to age, IQ, sex, or digit span performance. A Pearson  $r$  of .45 ( $p < .01$ ) was obtained, how-

TABLE 1

ANALYSIS OF VARIANCE OF TOTAL ATTITUDE SCORES

Source	SS	df	MS	F
Fantasy	.32	1	.32	3.01
Treatments	.84	1	.84	8.00**
Interaction	.27	1	.27	2.55
Within	7.96	76	.11	

\*\*  $p < .01$ .

ever, between the Fantasy Inventory and the CMA scale scores.

A two-way classification analysis of variance was computed for the total attitude score, obtained in the second session, to evaluate overall effects of treatment (experimental-control), fantasy level (High-Low), and the Fantasy  $\times$  Treatment interaction, which was the test of the main hypothesis. In this analysis, the data were found to violate the assumption of homogeneity of variance, and a logarithmic type transformation was utilized,  $\log(1 + X)$ , which eliminated the heterogeneity (Edwards, 1950).

Although the experimental production of frustration increased aggressive responses, there were no significant differences between levels of fantasy or between interaction of fantasy and experimental condition in total attitude scores as had been predicted (Table 1). A series of  $t$  tests does, however, show the expected significant differences on two of the four categories of the attitude questionnaire in the experimental condition (Table 2). The HF group was significantly more positive (lower score indicates more positive attitude) than the LF group in attitude toward the research ( $p < .05$ ), and the impersonal aspects of the Johnstone Center ( $p < .05$ ). There were no other significant differences, although both Johnstone Personnel and Examiner were in the predicted direction.

The differences between pre- and postfrustration Digit Span scores, and the Picture Technique data were also subjected to an analysis of variance; however, no further significant differences were obtained between fantasy or treatment groups. An  $F$  slightly short of the .05 level was obtained for treatments (experimental-control) with the Digit Span data.

In light of the significant correlation between fantasy scores and manifest anxiety, analyses were run to ascertain if performance on the Attitude Questionnaire, Digit Span test, and Picture Technique was related to performance on the CMA scale. There were no significant correlations between the CMA scale and any of these dependent variable measures.

## DISCUSSION

On the total Attitude Questionnaire the control group manifested much more positive feeling than the experimental group. The LF group also showed significantly more negative feeling regarding the research, and the impersonal elements of the Johnstone Center, than the HF group. The fact that the LF group did not express even more negativistic feelings than they did toward the examiner or the personnel of the Johnstone Center in light of findings showing retardates to be less guarded in expressing aggression (Beier et al., 1951), may be related to the long negative conditioning process to which these students have been subjected regarding any direct expression of aggression toward authority figures in the institution.

There were no significant differences obtained with the Digit Span test at the required level of confidence. However, the difference between the control and experimental treatment groups was only slightly short of the .05 level, and suggests that further investigation of this type performance measure may be desirable. The relatively restricted range of intellectual ability in this population may also have minimized the possibility of obtaining significant differences with this test.

The failure of the Picture Technique to yield any significant differences may be attributable to several factors; i.e., the underlying rationale of the technique may be faulty and esthetic determinants affecting the ranking of the sketches may be more, or as important, as the variables introduced in the experiment.

The findings of this study suggest that fantasy may be of some value in reducing aggressive, negativistic feelings, arising from situational frustrations. However, it appears that fantasy is not

TABLE 2

MEANS AND  $t$  VALUES FOR THE HIGH AND LOW FANTASY GROUPS ON THE SUBCATEGORIES OF THE ATTITUDE QUESTIONNAIRE FOR THE EXPERIMENTAL (Frustration) CONDITION

Attitude category	Group	$M$	$t$
Research	HF	2.14	2.43*
	LF	3.05	
Impersonal aspects, Johnstone	HF	1.91	2.35*
	LF	2.92	
Johnstone personnel	HF	1.14	1.98
	LF	1.90	
Examiner	HF	1.38	.97
	LF	1.66	

\*  $p < .05$ .



effective in reducing the basic anxiety levels measured by the *CMA* scale, as indicated by the positive correlation between these two variables. These two findings may help in the identification of the mechanisms responsible for the paradox that fantasy (though apparently characteristic of maladjustment) is helpful at times in tension reduction.

SUMMARY

This research was designed to test the hypothesis that HF subjects would manifest better post-frustration adjustment scores than the LF subjects. Institutionalized male and female high grade retardates were tested in two sessions. In the first session an estimate of the fantasy level and information regarding prefrustration level of other pertinent variables were obtained. The second session involved an experimental introduction of frustration, and the administration of various postfrustration adjustment measures. Significant results in the predicted direction were obtained with subscales of one of the three criterion measures. This finding, combined with a positive correlation between fantasy and manifest anxiety suggests that fantasy may function successfully as a defense mechanism under day-to-day, transient situational frustrations. However, it may be ineffective regarding the more chronic anxiety levels, and possibly here plays a detrimental role.

REFERENCES

BEIER, E. G., GORLOW, L., & STACEY, C. L. The fantasy life of the mental defective. *Amer. J. ment. Defic.*, 1951, 55, 582-589.

BELLAK, L. The concept of projection: An experimental investigation and study of the concept. *Psychiatry*, 1944, 7, 353-370.

CASTANEDA, A., MCCANDLESS, B. R., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, 27, 317-326.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

FESHBACH, S. Drive reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, 50, 3-11.

HACKBUSCH, F., & KLOPPER, B. The contribution of projective techniques to their understanding and treatment of children psychometrically diagnosed as feebleminded. *Amer. J. ment. Defic.*, 1946, 51, 15-34.

KAGAN, J. The measurement of overt aggression from fantasy. *J. abnorm. soc. Psychol.*, 1956, 52, 390-393.

LESSER, G. The relationship between overt and fantasy aggression as a function of maternal response to aggression. *J. abnorm. soc. Psychol.*, 1957, 55, 218-221.

PAGE, H. A. Studies in fantasy: Daydreaming and the TAT. *Amer. Psychologist*, 1956, 11, 392. (Abstract)

PAGE, H. A. Studies in fantasy—Daydreaming frequency and Rorschach scoring categories. *J. consult. Psychol.*, 1957, 21, 111-114.

SARASON, S. Projective techniques in mental deficiency. *Character Pers.*, 1945, 13, 237-245.

SEARS, R. R. Success and failure. In Q. McNemar & M. A. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942.

SHAFFER, L. F., & SHOBEN, E. J., JR. *The psychology of adjustment*. Boston: Houghton Mifflin, 1956.

(Received January 3, 1961)

# Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

## CONSULTING EDITORS

JOSEPH B. ADELSON  
*University of Michigan*

GORDON W. ALLPORT  
*Harvard University*

NORMAN H. ANDERSON  
*University of California, Los Angeles*

LEONARD BERKOWITZ  
*University of Wisconsin*

JACK W. BREHM  
*Duke University*

URIE BRONFENBRENNER  
*Cornell University*

DONALD R. BROWN  
*Bryn Mawr College*

DONALD T. CAMPBELL  
*Northwestern University*

LAUNOR F. CARTER  
*System Development Corporation*

DWIGHT W. CHAPMAN, JR.  
*Vassar College*

MORTON DEUTSCH  
*Bell Telephone Laboratories*

SEYMOUR FESHBACH  
*University of Pennsylvania*

DONALD W. FISKE  
*University of Chicago*

NATHANIEL L. GAGE  
*University of Illinois*

NICHOLAS HOBBS  
*Peabody College for Teachers*

MURRAY HORWITZ  
*New York University*

IRVING L. JANIS  
*Yale University*

HAROLD H. KELLEY  
*University of California, Los Angeles*

HERBERT C. KELMAN  
*Harvard University*

DANIEL J. LEVINSON  
*Harvard Medical School*

DAVID C. MCCLELLAND  
*Harvard University*

ROBERT B. MACLEOD  
*Cornell University*

HAROLD L. RAUSH  
*National Institute of Mental Health*

NEVITT SANFORD  
*Stanford University*

IRVING SARNOFF  
*Western Reserve University*

WILLIAM A. SCOTT  
*University of Colorado*

JOSEPH VEROFF  
*University of Michigan*

JUSTIN L. WEISS  
*Massachusetts Mental Health Center*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.



The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is \$10.00, or \$20.00 annually (foreign, \$20.50). Single issues are \$2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ  
University of Michigan  
Box 1248  
Ann Arbor, Michigan

---

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR  
Acting Production Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 Sixteenth Street, N.W.  
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
Prince and Lemon Sts., Lancaster, Pa.  
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Lancaster, Pa., and at additional mailing offices

Copyright © 1962 by the American Psychological Association, Inc.

## OWN ATTITUDE AS AN ANCHOR IN EQUAL- APPEARING INTERVALS<sup>1</sup>

HARRY S. UPSHAW

University of North Carolina

In recent years evidence has accumulated against Thurstone's (Thurstone & Chave, 1929) assumption that ratings of attitude statements in the method of equal-appearing intervals are independent of the attitudes of judges. Hovland and Sherif (1952) replicated Hinckley's (1932) study, arguing that the earlier finding of no effect on scale values due to judges' attitudes was traceable to Hinckley's use of a "carelessness criterion," which served to purge the judging group of people with extreme attitudes. Without the use of the disputed carelessness criterion, these investigators found that extreme judges tended to assign a large number of items to the categories at the end of the scale opposite to their own positions. Furthermore, a negative relationship was observed between the own positions of judges and the scale values assigned by them to midscale items. These results were interpreted as support for three hypotheses:

*Hypothesis I.* Judges with extremely pro or con attitudes will show a tendency to concentrate their placement of items into a small number of categories.

<sup>1</sup> This study was supported by a PHS grant (No. GN-5549) from the Division of Research Grants, United States Public Health Service.

The author is grateful to other members of the project staff of the "Study of the Essentials for Public Health Nursing Field Experience" for their aid in various phases of the research. In addition, appreciation is expressed to the following people for their particular help: John W. Thibaut, Peter Gumpert, and Bibb Latané, for stimulating discussions that led to the formulation of the problem and experimental design; W. J. Daniel and Paul Wood, for making arrangements for subjects; E. D. Hinckley, for making available the items used in his earlier study; Charles D. Ward, for numerous suggestions and other help in the analysis of data; Thomas Ostrom and Samuel Fillenbaum, for suggestions concerning the preparation of the manuscript. Finally, gratitude is expressed to the reviewer for this journal who made many suggestions for improvement of the report.

*Hypothesis II.* Judges with an extreme position and strong involvement will be highly discriminating in accepting items at their own end of the scale. They will correspondingly display a strong tendency to lump together statements at the end of the scale which they reject. The former tendency can be described as a raised *threshold of acceptance* and the latter as a lowered *threshold for rejection*.

*Hypothesis III.* A greater degree of displacement will occur for the "neutral" items and a smaller degree for the sharply defined pro and con statements at the extremes (Hovland & Sherif, 1952, p. 824).

The general flavor of the Hovland and Sherif report, and of Hypotheses II and III in particular, suggests an interpretation of the results in terms of a process of perceptual vigilance, whereby judges displace neutral items to the end of the scale opposite to their own position.

In a later paper, Hovland, Harvey, and Sherif (1957) presented a model to account for the effect of judges' attitudes on the evaluation of the position represented in a communication. Manis (1960) employed this model to explain the relationship he observed between the two variables. According to the model, which will be called here the "assimilation-contrast" model, items that a subject would personally reject are judged to be more distant from his own position than they really are (contrast), and items that represent positions only slightly discrepant from a subject's own attitude are judged to be closer to his own position than they really are (assimilation). It is interesting to note that the assimilation-contrast model makes allowance for both positive and negative relationships between judges' attitudes and item scale values. As a matter of fact, Manis observed a positive relationship for midscale items, whereas Hovland and Sherif found a negative relationship.



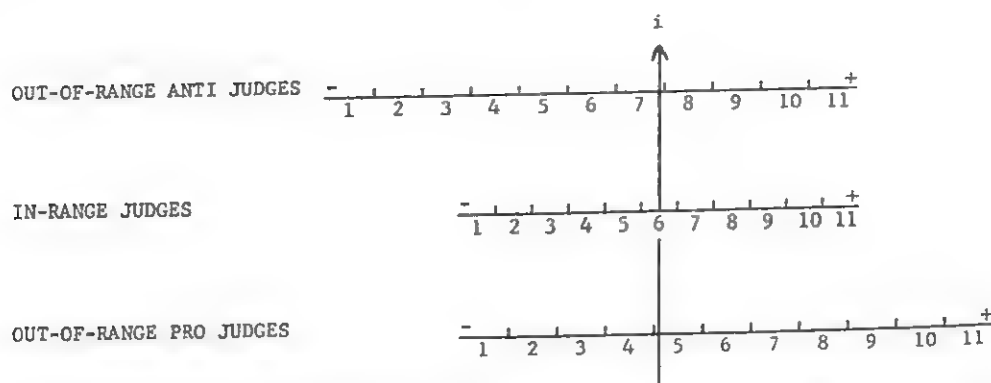


FIG. 1. Hypothetical scales adopted by out-of-range and in-range judges showing differences in judgments of an item  $i$  as a function of differences in these scales.

Still another theoretical orientation to the relationship between judges' attitudes and scale value is represented by the research of Segall (1959), who approached the problem as one of adaptation level. According to this model, a judge's attitude, in part, determines which stimulus he judges to be neutral, causing the point of neutrality, or adaptation level, to be displaced toward the judge's attitude. Because all absolute judgments are assumed to be based on the perceived distance of each item from the center of the reference scale, the judge's own position affects all of his judgments. Thus, a negative relationship between own position and scale value is hypothesized. Segall did not observe any systematic relationship between these variables, presumably because of an inadequate range of attitudes among his judges. Despite this failure to confirm the adaptation level hypothesis with respect to judges' attitudes, the model remains an important one to consider in the context of the equal-appearing intervals judging task.

A fourth theoretical model, which may be called "variable series," also applies to the effect of judges' attitudes on scale values. This model is based on the theoretical position of Volkmann (1951), who assumes that the stimulus range is the major determinant of the particular values assigned to items in a series. Judges whose own attitudes are included in the range of statements presented for judgment (i.e., *in-range* judges) are expected to divide the item range into segments corresponding to the categories specified in the experimental instructions. It may be assumed that the equal-appearing intervals

judging task makes salient to a judge his own attitude, which functions for *out-of-range* judges as an anchor at one end of the scale. Thus, out-of-range judges subjectively divide a range corresponding to the experimental series augmented by their own attitudes. Accordingly, the scales adopted by in-range judges are narrower than those of out-of-range judges, leading to relatively finer discriminations among the items by in-range judges. Figure 1 illustrates the manner in which variable series accounts for a predicted negative relationship between own position and item judgments for an item that is in the middle of the series, and consequently at the middle of the scale adopted by in-range judges. The differences in item judgment illustrated in Figure 1 would be obliterated by extending the item range to include the positions of all judges, thus creating a constant series for all subjects.

The present experiment attempts to test differentially four theoretical interpretations of the relationship between judges' attitudes and equal-appearing intervals scale values. Inherent in the models of perceptual vigilance, assimilation-contrast, adaptation level, and variable series are two critical independent variables relevant to this relationship: the range of items presented for judgment, and the own positions of judges. The present design includes manipulation of both of these variables. Three experimental conditions were used:  $t$  (total series), in which it is assumed that the own positions of all judges fall within the item range,  $a+$  (series aborted at the pro end) and  $a-$  (series aborted at the anti end), in which it is assumed that the own

positions of many judges exceed the range. Within each of the experimental conditions, scale values of the items included in the series were determined for three groups of judges: A (anti judges), N (neutral judges), and P (pro judges). Nine sets of equal-appearing intervals item values were computed for the three groups of judges in the three experimental conditions. In the discussion to follow, the subgroups of judges and their sets of scale values are designated by a letter corresponding to the judges' attitudes, followed by a subscript representing an experimental condition (e.g.,  $A_1$  represents anti judges in the total series condition,  $N_a$  represents neutral judges in the condition in which the series was aborted at the pro end, etc.).

In terms of the present design, the four explanatory models are conceptualized as follows:

*Perceptual vigilance.* The critical independent variable in this model is personal involvement. Presumably, the more extreme a judge's own attitude, the greater his personal involvement, and the more likely he is to be threatened by the judging task. As a consequence of the threatening nature of the task, extremely pro and anti judges displace items in a direction opposite to their own attitudes. Neutral judges, having less personal involvement in the task, give more nearly accurate judgments than do extreme judges. Insofar as the absolute positions of the judges' attitudes is the factor accounting for any differences in scale values among the nine subgroups of the study, the manipulation of the item series has no relevance. With respect to the two independent variables of this study, perceptual vigilance predicts a main effect due to own attitude of judges, but no effect due to manipulation of the series, or to interaction. It is, however, assumed that the perceptual distortions of extreme judges are more likely to occur with ambiguous stimuli than with clear-cut, well-structured stimuli. Edwards (1946) has shown that midscale items on an equal-appearing intervals scale tend to be more ambiguous than extreme items. For this reason, it is expected that the predicted difference between judges,  $A > N > P$  (where ">" means "more favorable

than") for all experimental conditions, occurs primarily in the judgments of midscale items.

*Assimilation-contrast.* In explaining the influence of judges' attitudes on scale values, this model assumes that each judge displaces, toward his own position, items that are only slightly removed from those he would endorse, and displaces, away from his position, items that he would reject. It is not feasible within the present design to compute latitudes of acceptance and rejection for each judge. However, on the assumption that pro judges would reject the most anti items and that anti judges would reject the most pro items, the assimilation-contrast model implies a significant main effect due to the variable of judges' attitudes. Furthermore, if assimilation (displacement toward own position) is observed at all, it must be for items that are closer to own position than any items displaying the contrast phenomenon (displacement away from own position). Like the perceptual vigilance model, assimilation-contrast predicts no effect due to manipulation of the item series, and no effect due to interaction between judges' attitudes and item series.

*Adaptation level.* This model assumes that item judgments depend on the perceived distance of each item above or below the adaptation level (point of neutrality). A judge's adaptation level is the result of three types of influence: a weighted average of the stimulus values of items in the series; an effect due to anchor stimuli or other contextual factors; and a residual effect due to pre-experimental factors (Michels & Helson, 1949). The attitudes of judges may be conceived, as in Segall (1959), to contribute to a residual effect, or to serve an anchor function. With either conception, it is assumed that the adaptation level of each judge is displaced toward his own attitude, producing a displacement of item judgments in the opposite direction. A main effect due to judges' attitudes is predicted, in the order:  $A > N > P$ . Because it produces variation in the weighted average of the stimulus values, the manipulation of the item series leads to predicted shifts in adaptation level toward the anti end of the scale in the  $a +$  condition, and toward the pro end in the  $a -$  condition. As a result of these shifts in adaptation level, a difference due to



experimental conditions is predicted, with item values in the order:  $a+ > t > a-$ . Hence, unlike perceptual vigilance and assimilation-contrast, which predict a single main effect, adaptation level predicts main effects due to item series, as well as to judges' attitudes.

*Variable series.* The critical factor in this interpretation is whether the item series includes a judge's own position. In-range judges are assumed to adopt reference scales defined by the items at either extreme of the item series, whereas out-of-range judges adopt scales anchored by their own attitudes at one end, and the most extreme item in the series at the other. It is intended that the series in the  $t$  condition covers the range of judges' attitudes. Therefore, the prediction is made that  $A_t = N_t = P_t$ . With the most favorable items removed in the  $a+$  condition, the scales adopted by  $P$  judges are predictably wider than those of  $N$  and  $A$  judges due to differential anchoring at the pro end. The result of the varying widths of reference scales is variation in item judgments as a function of own position, a given item receiving a lower value on a scale that extends farther at the pro end. Therefore, the predicted order of item judgments is  $A_{a+} \geq N_{a+} > P_{a+}$ . ( $A_{a+}$  is predicted to be equal to or greater than  $N_{a+}$ , depending on whether the experimentally aborted series covers the positions of both groups.) By similar reasoning for the  $a-$  condition, the predicted order of item judgments is  $A_{a-} > N_{a-} \geq P_{a-}$ , due to the extension of the scales adopted by  $A$  judges in the anti direction. Phrased another way, it is the manipulation of the series that produces a difference due to own positions. Thus, variable series predicts a main effect due to judges' attitudes, as do the perceptual vigilance, assimilation-contrast, and adaptation level models. It also predicts a main effect due to manipulation of the series, as does adaptation level. The distinctive prediction for variable series, however, is an interaction between own position and item series. In particular, the prediction is that, because  $A$ ,  $N$ , and  $P$  judges shift differentially in scale width as a function of the experimental conditions, the difference between the  $t$  and  $a+$  conditions is smaller for  $P$  judges than for  $A$

and  $N$  judges, and the difference between the  $t$  and  $a-$  conditions is smaller for  $A$  judges than for  $P$  and  $N$  judges.

## METHOD

### Procedure

In order to provide maximum comparability with previous research, the present study is essentially a replication of the Hinckley (1932) and Hovland and Sherif (1952) studies. The major difference between the present design and that of the earlier studies is that the range of the item series, as well as judges' attitudes, is treated as an independent variable. The manipulation of item range was accomplished, as in Fehrer (1952), by preparing three sets of attitude items. One set is identical to that employed by Hinckley and by Hovland and Sherif, consisting of all of the 114 statements assembled by Hinckley relating to the social position of the Negro. The second set of items represents an aborted range of statements with the 28 most favorable items removed on the basis of the original Hinckley scale values. (The decision concerning the number of items to be removed was an arbitrary one.) The third set has the 28 most unfavorable items removed.

Five hundred students of introductory psychology at the University of North Carolina served as judges, partitioned at random into three groups, each group receiving one of the sets of items. All judges received the following instructions, adapted from Hinckley (1932):

1. You have been solicited as a subject in the construction of a scale for measuring attitude toward the Negro.
2. This envelope contains a series of opinions about the *social position of the Negro*. Your task is to sort these opinions into 11 piles.
3. To aid you in this task, you will find enclosed eleven slips with Roman numbers on them. Please lay these out in the following order: I, II, III, IV, V, VI, VII, VIII, IX, X, XI. *On slip I*, put those statements which place the Negro in the lowest social position. *On slip XI*, put those statements which place the Negro in the highest social position. In like manner, place the other statements on slips II, III, IV, etc., as they give the Negro a higher and higher social position. You will then have all of the opinions arranged in eleven piles, corresponding to eleven steps in the social ladder.
4. Do not try to get the same number in each pile. They are not evenly distributed.
5. You can probably finish the task in less than an hour.
6. Kindly place your name and classification on enclosed slip.
7. When you have finished, please clip all statements which you have placed in each pile together, with the proper Roman number slip on top. Replace the eleven piles thus clipped into the envelope.

After the sorting task, all judges completed the Murphy-Likert Scale of Attitude toward the Negro (Likert, 1932). On the basis of the distribution of their scores, three groups were constituted, with those in the most unfavorable quintile forming the A group, those in the middle quintile forming the N group, and those in the most favorable quintile forming the P group. The Murphy-Likert scale has a maximum score range of 20-70. The actual range of scores in this study was 38-70. The anti-Negro quintile was defined in terms of scores lower than 55; the neutral quintile, scores between 59 and 62; and the pro-Negro quintile, scores higher than 65. Only A, N, and P judges were of concern in this study. After eliminating judges in the second and fourth Likert quintiles, 321 subjects remained.

### Data Analysis

Nine sets of equal-appearing intervals values were computed for the various subgroups corresponding to judges' attitude and the range of items judged. The subgroups varied in size from 29 in the A<sub>1</sub> condition to 45 in the P<sub>1</sub> condition. "Control scale values" were computed, based on the judgments of the entire group of 175 subjects who participated in the t condition, including those in the second and fourth Likert quintiles, as well as those in conditions A<sub>1</sub>, N<sub>1</sub>, and P<sub>1</sub>. On the basis of the control scale values, the nine sets of values for the experimental subgroups were partitioned into subsets of extremely pro items (13 items with scale values between 9.50 and 11.49), moderately pro (18 items with values between 7.50 and 9.49), midscale (30 items between 4.50 and 7.49), moderately anti (24 items between 2.50 and 4.49), and extremely anti items (29 items between 0.50 and 2.49).<sup>2</sup>

The hypotheses for this study, derived from the four theoretical models, are phrased in terms of main effects and interaction. This formulation suggests a two-way analysis of variance as an appropriate statistical procedure. However, it should be noted that one of the four models, variable series, is based upon consideration of the effect of the two independent variables on the width of the reference scales adopted by judges. Whereas variations in scale width lead to predictable differences in the modal judgments of items, such variations also lead to predictable effects on the dispersion of judgments, which may render untenable the homogeneity of variance assumptions of the analysis of variance statistical models. As a matter of fact, preliminary analyses of some of the

present data by means of the method of successive intervals, which permits solution of the item discriminational dispersions, indicate that there are significant effects on item variance due to judges' attitudes, item series, and interaction between these variables.<sup>3</sup> To overcome this difficulty, a nonparametric analogue of the analysis of variance was used, employing a statistic called  $\chi^2_r$ , the distribution of which approaches that of  $\chi^2$  as the number of replications (in this case, items) becomes indefinitely large (Friedman, 1937, 1940; Wilcoxon, 1949).

In order to test the main effect due to judges' attitudes, it was necessary to determine for A, N, and P judges scale values that were not systematically influenced by the manipulation of item series. To this end, the medians of the values assigned to each item under the three experimental conditions were computed for the three groups of judges. On the basis of these medians the three judging groups were ranked, item-by-item, for extremely pro, moderately pro, midscale, moderately anti, and extremely anti items. For each of the item subsets, a sum of ranks was computed, and these sums for A, N, and P judges were tested for equality by means of  $\chi^2_r$ . An over-all test of the significance of own attitude was made by summing the values of  $\chi^2_r$  and the degrees of freedom for the five item subsets. The evaluation of the effect of the item series was conducted in a manipulation of item series, the median scale values assigned in the t condition by A, N, and P judges were determined, as were the medians for the a+ and a- conditions. The experimental conditions were ranked, item-by-item, for each of the subsets of items. Again,  $\chi^2_r$  was used to test the hypothesis that the sum of ranks assigned to the experimental conditions are equal. Summing the  $\chi^2_r$ 's over the item subsets provided an over-all test of the effect due to item series.

There are two components of the interaction effect predicted from the variable series model. First, the differences between judgments in the a+ and t conditions were expected to be greater for A and N judges than for P judges. This expectation is based upon consideration of the relatively narrower scale of A and N judges in the a+ condition, with an accompanying displacement of judgments in the direction of the aborted end of the scale. To test this hypothesis, the scale values assigned by A, N, and P judges in the t condition were subtracted, item-by-item, from those obtained in the a+ condition. The A, N, and P judges were ranked according to these differences, the rank sums were evaluated by  $\chi^2_r$  for each item subset, and an over-all test was made on the basis of the sum of the  $\chi^2_r$ 's for the five item subsets. Similarly, the differences between judgments in the a- and t conditions were expected to be greater for N and P judges than for A judges, due to the maintenance of a relatively wide scale by the latter in this aborted series condition. As a result of their narrower scales in the a- condition, N and P judges were

<sup>2</sup> An 11-page document listing the items constituting the three experimental series, the obtained scale and Q values for the nine subgroups, and the control scale and Q values is available through the American Documentation Institute. Order Document No. 6942 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>3</sup> The data will be the subject of a later report.



TABLE 1  
MEAN ITEM VALUES

Experimental condition	Attitude of judges	Item subset				
		Extremely pro	Moderately pro	Mid-scale	Moderately anti	Extremely anti
t	A	9.87	8.22	6.15	3.98	2.09
	N	10.20	8.56	6.23	3.26	1.62
	P	10.33	9.03	5.87	3.19	1.53
a +	A	—	—	6.64	4.33	2.46
	N	—	—	6.63	3.73	1.79
	P	—	—	5.83	2.93	1.40
a —	A	9.79	8.27	6.61	3.75	—
	N	9.72	8.30	5.89	2.93	—
	P	9.98	8.61	5.40	2.53	—
Control		10.14	8.63	6.04	3.37	1.66

Note.—A dash indicates that an experimental manipulation rendered impossible the collection of data otherwise appropriate for a cell in the table. Dashes appear for both extremely pro and moderately pro items in the a + condition and for only the extremely anti items in the a — condition because of the generally lower values obtained in the present study as compared with those obtained by Hinckley (1932). Only four of the anti items removed in the a — condition, on the basis of the Hinckley data, received moderately anti values in the present study.

expected to displace their judgments toward the aborted, anti end of the scale. Accordingly, scale values obtained for each item in the t condition were subtracted from the corresponding values in the a — condition, and the differences evaluated by the ranking procedure for A, N, and P judges. The  $\chi^2$ 's computed for each item subset were summed to obtain an over-all evaluation of differences between a — and t conditions. Finally, the  $\chi^2$ 's corresponding to the two components of the interaction, a — minus t and a + minus t, were summed to provide a means of testing the total effect due to interaction.

### RESULTS

Table 1 presents the mean scale values assigned to items of each subset by A, N, and P judges under the three experimental conditions. Table 2 summarizes the results

of the nonparametric tests of significance of the effects due to own attitude and item series, and Table 3 summarizes the results of the tests of interaction. The relative sizes of the rank sums listed in Tables 2 and 3 reflect the order of scale values assigned by the judges participating in the designated comparisons, with lower sums generally corresponding to lower scale values. From Tables 2 and 3 it can be seen that, considering all items, the effects due to own attitude, item series, and interaction are all highly significant. Of the four explanatory models, only variable series accounts for all three of these effects. However, not all of the analyses for separate item subsets are consistent with this

TABLE 2  
SUMS OF RANKS AND TESTS OF SIGNIFICANCE FOR MAIN EFFECTS

Item subset	Attitude of judges			$\chi^2$	df	Experimental condition			$\chi^2$	df
	A	N	P			t	a +	a —		
Extremely pro	19.0	23.0	36.0	12.15***	2	23.5	—	15.5	4.92*	1
Moderately pro	30.0	31.0	47.0	10.11**	2	31.5	—	22.5	4.50*	1
Midscale	65.0	67.0	42.0	13.31***	2	62.5	69.5	42.0	14.09***	2
Moderately anti	71.0	47.0	26.0	42.25***	2	48.0	66.0	30.0	27.00***	1
Extremely anti	81.0	52.0	29.0	50.30***	2	31.5	49.5	—	12.00***	1
All items				128.12***	10				62.51***	7

Note.—A dash indicates that an experimental manipulation rendered impossible the collection of data otherwise appropriate for a cell in the table.  
\*  $p < .050$ , two-tailed test.  
\*\*  $p < .010$ , two-tailed test.  
\*\*\*  $p < .005$ , two-tailed test.

TABLE 3  
SUMS OF RANKS AND TESTS OF SIGNIFICANCE FOR INTERACTION

Component of interaction	Item subset	Attitude of judges			$\chi^2$	df
		A	N	P		
a — minus t	Extremely pro	32.0	23.0	23.0	4.15*	2
	Moderately pro	43.5	34.5	30.0	5.25**	2
	Midscale	77.0	51.0	46.0	19.10****	2
	Moderately anti	51.5	52.0	40.5	3.52*	2
	Extremely anti	—	—	—	—	—
	All items	—	—	—	32.02****	8
a + minus t	Extremely pro	—	—	—	—	—
	Moderately pro	—	—	—	7.66***	2
	Midscale	69.0	57.0	48.0	20.58****	2
	Moderately anti	51.0	62.0	31.0	16.22****	2
	Extremely anti	64.0	61.0	37.0	—	—
	All items	—	—	—	44.46****	6
Combined	Extremely pro	—	—	—	4.15*	2
	Moderately pro	—	—	—	5.25**	2
	Midscale	—	—	—	26.76****	4
	Moderately anti	—	—	—	24.10****	4
	Extremely anti	—	—	—	16.22****	2
	All items	—	—	—	76.48****	14

Note.—A dash indicates that an experimental manipulation rendered impossible the collection of data otherwise appropriate for a cell in the table.  
 \*  $p < .100$ , one-tailed test.  
 \*\*  $p < .050$ , one-tailed test.  
 \*\*\*  $p < .025$ , one-tailed test.  
 \*\*\*\*  $p < .001$ , one-tailed test.

conclusion. For example, although a significant difference was found among A, N, and P judges for moderately pro and extremely pro items, the direction of the difference is opposite to that predicted not only by variable series, but also by perceptual vigilance and adaptation level. The finding that higher (more pro) values were assigned to these items by pro judges is consistent with the assimilation-contrast model. However, for two reasons, caution must be exercised in claiming support for the model on the basis of this finding. First, a contrary result, that pro judges assigned lower values to these items, would have been equally consistent with the model (interpretable as contrast, rather than assimilation). Second, the significant effects due to item series and interaction observed for these items were not anticipated on the basis of the model. Furthermore, as will be discussed in the next section, there are supplementary data suggesting that some of the items showing the apparent assimilation effect are well within

the latitudes of acceptance of P judges, whereas the Hovland et al. (1957) model specifies assimilation of items somewhat removed from this region of the scale.

The analyses of the effect of own position on judgments of midscale, moderately anti, and extremely anti items are consistent with the predictions from all four models (interpretable as contrast effects in terms of the assimilation-contrast model).

With respect to the experimental manipulation of item range, the results reported in Table 2 are consistent with the predictions from both adaptation level and variable series for each of the item subsets. The demonstration of a significant effect due to the item series is not, in itself, relevant to scale construction by the method of equal-appearing intervals, which conventionally involves a standard series for all judges. If, as conceptualized by the variable series model, the attitudes of out-of-range judges contribute to the definition of their reference scales, the item series assumes importance by virtue of



the interaction between it and own attitudes.

Variable series is the only model of the four considered in this research which specifies an interaction between the two independent variables. It, furthermore, specifies the direction of the interaction. Therefore, the data reported in Table 3 constitute a critical test of that model. By reference to the table, it can be seen that, over-all, the interaction effect is highly significant. The obtained  $\chi^2_r$ 's for the  $a -$  minus  $t$  component for extremely pro and moderately anti items fall short of the values conventionally adjudged statistically significant. In both of these cases, however, the direction of the differences in rank sum for A, N, and P judges is that predicted by variable series. (It should be noted that, while the  $\chi^2_r$  for extremely pro items in the combined interaction also falls short of the acceptable level, this is due entirely to the  $a -$  minus  $t$  component, since the combined term for extremely pro items was comprised only of that component.) Whether or not allowances are made for the two discrepant cases, collectively the analyses reported in Table 3 provide evidence in support of the variable series prediction that judges' attitudes and the item series interact in their influence on equal-appearing intervals scale values. This conclusion can be drawn even though it is apparent in the analyses of the main effect due to judges' attitudes that the variable series model cannot account for the data obtained with extremely pro and moderately pro items.

#### SUPPLEMENTARY DATA

The variable series model was formulated on the basis of evidence in the psychophysical literature that a stimulus not present in the series can serve an end anchor function, defining the upper or lower end of the reference scale (e.g., Hunt & Volkman, 1937). As envisaged in the present design, the total series would contain items more extreme at either end than any judge's own attitude. With the removal of the most pro items for the  $a +$  condition, it was assumed that the upper end anchors for pro judges would be their own attitudes, whereas the upper end anchors for neutral and anti judges would be either their own attitudes or the most favor-

able item in the aborted series. In any event, it seemed reasonable to assume that the reference scales adopted by A, N, and P judges would, because of differential end anchoring, differ in width more in the  $a +$  condition than in the  $t$  condition. Furthermore, it was assumed that there would be greater similarity between the scales of  $P_t$  and  $P_a$  judges than between those of either  $A_t$  and  $A_a$  judges, or  $N_t$  and  $N_a$  judges. By the same reasoning, greater similarity was assumed between the scales of  $A_t$  and  $A_a$  judges than between either  $P_t$  and  $P_a$  judges, or  $N_t$  and  $N_a$  judges.

After the decision to use the Hinckley (1932) items in this research, it was deemed inadvisable to assess own position by means of the same items. Memory of the rating previously assigned to an item could, conceivably, influence the likelihood of a subject's endorsing the item. Conversely, memory of an endorsement or rejection of an item could alter the rating of the item. Perhaps it would have been wise to have removed from the item pool a subset of items with which own position could have been measured. It was decided, however, to assess the attitudes of judges by means of the Murphy-Likert scale. Likert scales are ordinal scales, permitting only an ordering of subjects according to their attitudes. This type of measure is adequate in the present design for the two aborted range conditions. However, minimal validity of the experimental manipulations for testing variable series depends upon two assumptions. First, stimulus anchoring (as opposed to attitude anchoring) is assumed in the  $t$  condition at the pro end for anti judges, and at the anti end for pro judges. Second, attitude anchoring is assumed at the pro end for pro judges in the  $a +$  condition, and at the anti end for anti judges in the  $a -$  condition.

To assess the validity of the experimental manipulations for providing a test of the variable series model, new data were collected. A questionnaire was administered to 201 students drawn from the same undergraduate psychology course that a year earlier supplied the judges in the main experiment. The questionnaire included a 20-item equal-appearing intervals scale based on the control values obtained in the main experiment. The

items were chosen to obtain maximum coverage of the attitude continuum with minimum *Q* values. The scale appeared as the third part of the questionnaire, the first two parts being concerned with the judgment of items, none of which were included in the 20-item attitude scale. The following instructions were given, adapted from those used by Hovland et al. (1957) to determine latitudes of acceptance and rejection:

1. This part contains 20 statements concerning the social position of the Negro. This time, however, you are requested to indicate your own acceptance or rejection of the items as expressions of your own opinion.
2. Read all of the items carefully before you make any judgments.
3. When you have read all of the statements look at the table on the next page. This table has five headings: Most Acceptable, Also Acceptable, Neither Acceptable nor Objectionable, Also Objectionable, and Most Objectionable.
4. Your task is to:
  - a. Find the *one* item in this set of 20 items that is *most acceptable* from your point of view, and list the item number for that one statement under the Most Acceptable column in the table.
  - b. There may be other statements which you find *also acceptable* from your point of view. Put the item numbers for these statements (if there are any) under the Also Acceptable column of the table.
  - c. Now find the *one* statement which is *most objectionable* from your point of view, and enter that statement's number under the Most Objectionable column in the table.
  - d. There may be other statements which you find *also objectionable* from your point of view. List the numbers of these statements (if there are any) under the Also Objectionable column of the table.
  - e. In the last column, Neither Acceptable nor Objectionable, the numbers of the remaining items (if there are any) should be listed. These remaining statements should be statements that from your point of view are *neither acceptable nor objectionable*.
  - f. As a check to make sure that all 20 items have been listed in the table, count the item numbers listed there. If not all 20 have been listed, find the missing items and list their numbers in the appropriate columns.
  5. As in previous parts, the items have identification numbers, but these, again, should be ignored.

The fourth part of the questionnaire was the Murphy-Likert scale, on the basis of which A, N, and P judges were selected by the same criteria used in the main experiment. Of the 201 judges, 47 were thus assigned to the A group, 45 to the N group, and 37 to the

P group. For each judge in the three groups the range of acceptable items on the equal-appearing intervals scale was determined (whether recorded as "most acceptable" or "also acceptable"). Table 4 shows the lower and upper boundaries of acceptable items for the three groups. These boundaries are represented in the table by the cumulative frequency and cumulative proportion of judges in each group for whom any item of the scale is the lowest and highest valued item endorsed. It is interesting to note that, considering the most extremely anti items endorsed, for only 16 (a proportion of .34) of the A judges were these items in the region that was excluded in the a- condition. Approximately 25% of the items in the Hinckley (1932) series are in this region. Furthermore, more than one-fourth of the A judges endorsed items with values above 9.69. This information, together with the finding that all 37 of the P judges endorsed the most extremely pro item of the scale, suggests that the total Hinckley series does not adequately cover the range of attitudes of white judges selected from this southern university. If this is so, it must certainly be true that the Negro judges in the Hovland and Sherif (1952) study were all out-of-range judges.

Assuming that the student population that supplied data for Table 4 was the same as that participating in the main experiment, the table provides information on the basis of which the validity of the earlier experimental manipulations can be evaluated. If those manipulations were minimally adequate to provide a test of the variable series formulation, stimulus anchoring must have prevailed in the t condition for A judges at the pro end, and for P judges at the anti end. It is likely that the large number of A judges who endorsed the most pro item of the scale reflects an end effect, and that many of these people would have endorsed items that are even higher in scale value, if they had been offered. Although it appears likely that stimulus anchoring did not prevail in this situation for all judges, it is probable that it did prevail for perhaps as many as three-fourths of the A group. At the anti end, it seems reasonable to conclude that stimulus anchoring prevailed for all P judges.



TABLE 4

CUMULATIVE FREQUENCY AND CUMULATIVE PROPORTION OF MOST EXTREMELY ANTI AND MOST EXTREMELY PRO ITEMS ENDORSED BY A, N, AND P JUDGES

Item scale value	Most anti item endorsed						Most pro item endorsed					
	A judges		N judges		P judges		A judges		N judges		P judges	
	Cumulative frequency	Cumulative proportion	Cumulative frequency	Cumulative proportion	Cumulative frequency	Cumulative proportion	Cumulative frequency	Cumulative proportion	Cumulative frequency	Cumulative proportion	Cumulative frequency	Cumulative proportion
10.52	47	1.00	45	1.00	37	1.00	47	1.00	45	1.00	37	1.00
10.18	47	1.00	45	1.00	37	1.00	36	.77	6	.13	0	.00
9.69	47	1.00	45	1.00	37	1.00	35	.74	5	.11	0	.00
9.23	47	1.00	45	1.00	37	1.00	10	.21	0	.00	0	.00
8.53	47	1.00	45	1.00	37	1.00	9	.19	0	.00	0	.00
8.33	47	1.00	44	.98	37	1.00	8	.17	0	.00	0	.00
7.57	47	1.00	43	.96	33	.89	6	.13	0	.00	0	.00
7.27	47	1.00	41	.91	23	.62	6	.13	0	.00	0	.00
6.86	46	.98	37	.82	18	.49	6	.13	0	.00	0	.00
5.98	46	.98	37	.82	17	.46	4	.09	0	.00	0	.00
5.71	46	.98	35	.78	13	.35	3	.06	0	.00	0	.00
5.18	45	.96	32	.71	12	.32	1	.02	0	.00	0	.00
4.56	44	.94	32	.71	12	.32	0	.00	0	.00	0	.00
4.34	39	.83	21	.47	6	.16	0	.00	0	.00	0	.00
3.65	32	.68	14	.31	3	.08	0	.00	0	.00	0	.00
3.15	30	.64	14	.31	3	.08	0	.00	0	.00	0	.00
2.55	18	.38	7	.16	3	.08	0	.00	0	.00	0	.00
2.16	16	.34	6	.13	3	.08	0	.00	0	.00	0	.00
1.63	9	.19	0	.00	1	.03	0	.00	0	.00	0	.00
1.14	4	.09	0	.00	0	.00	0	.00	0	.00	0	.00

Note.—Items in the first segment of the table were excluded in the a + condition. Those in the third segment were excluded in the a - condition.

A second assumption on which the validity of the manipulations depends is that attitude anchoring prevailed at the pro end for P<sub>n</sub> judges, and at the anti end for A<sub>n</sub> judges. Table 4 indicates that all P judges endorsed statements above the pro stimulus anchor of the a + series and that .34 of the A judges endorsed statements below the anti stimulus anchor of the a - series. Again it can be concluded that the criterion was probably met by the P judges. The criterion with respect to A judges was probably met by about one-third of the judges.

On the basis of the data relating to the two critical assumptions, it seems very likely that the experimental procedure was adequate for P judges. There are grounds for disputing the validity of the procedure for A judges, although it is probable that enough of these judges satisfied the criteria to account for the results obtained with A judges in the main experiment.

In the presentation of results of the main experiment it was noted that P judges assigned higher values to extremely pro and moderately pro items than did A judges. In terms of the assimilation-contrast model, this finding could represent assimilation on the part of P judges. Assimilation is assumed to occur at, or slightly beyond, the boundary of a judge's latitude of acceptance. The procedures used by Hovland et al. (1957) to determine latitudes of acceptance were employed in the collection of the data represented in Table 4. From the table it can be seen that all P judges endorsed the item with scale value 10.52, and that they all endorsed other items with values of 8.33, or lower. To the extent that the subjects supplying these supplementary data are like those participating in the main experiment, assimilation by P judges should occur only for items with values equal to, or lower than, 8.33. The extremely pro items, however, are items between 9.50 and 11.50, whereas moderately

pro items have values between 7.50 and 9.49. Hence, the extremely pro items are well within the latitude of acceptance of P judges. This fact, together with other evidence already discussed, casts suspicion on the adequacy of assimilation-contrast, as presently formulated, to account for the data with respect to pro items; data which none of the other models can account for, either.

#### DISCUSSION

The data of this study indicate that there is a differential effect of own position on equal-appearing intervals scale values according to the range of items presented for judgment. Of the four explanatory models considered, only variable series specifies such an interaction. However, the positive relationship between own position and scale value obtained in this study for pro items cannot be simply explained by variable series or any of the other models considered here. Perhaps own position is not the critical variable in this context, or perhaps it is not the only critical variable. One interpretation of the present experiment is that, when confronted with a judgmental task involving an item series which, for him, is aborted, a judge supplies anchors based on his expectation concerning what the series should be. The judgmental task may make salient to the judge various anchors at several points along the continuum, only one of which is his own attitude. If the series does not include items around these points, then, the self-supplied anchors may assume importance in determining judgmental behavior.

It may be objected that the choice of subjects did not permit an adequate test of the perceptual vigilance interpretation. Hovland and Sherif (1952) selected judges with extreme attitudes and known personal involvement in the race issue. The present study, on the other hand, used only judges of the type described by Hovland and Sherif as "average white judges." The "average white judges" in this study became extreme judges in some experimental conditions by virtue of the manipulation of the item series. Whether judges who are extreme in their "natural state" (presumably as defined by national norms) behave differently from those

called extreme in this study is a problem now being investigated by Thomas Ostrom of this university. Pending Ostrom's results, the influence of truly extreme attitudes on judgments of the type considered here is a matter of conjecture. However, any hypothesis concerning the judgmental processes of involved, extreme judges should take into account the interaction effect obtained in this study between own attitude and the item series. Perceptual vigilance as presently formulated does not.

The assimilation-contrast model, similarly, does not consider the influence of the item series. In the experiment accompanying the statement of the model (Hovland et al., 1957), the evaluations of single, 15-minute oral communications were analyzed in terms of the model. Manis (1960) analyzed the judgments of 12 attitude statements. It could be that the model is appropriate for some conditions of item series, but not for others. Research directed to these limits of applicability is suggested by the findings of this study.

Although the data concerning the two main effects due to own attitude and item series are consistent with the adaptation level model, those concerning the interaction between these variables are not. On the surface, it is difficult to imagine how any influence on the point of perceived neutrality can be made to account for this interaction.

Of the four explanatory models, variable series is best supported by the data. Applied to the judgmental task of equal-appearing intervals, the variable series interpretation assumes, in effect, that the own attitudes of judges are an extraneous variable which acquires importance only when the item series is such that the positions of some judges are outside it. A reference scale, defined by end anchors corresponding to the most pro and anti items, is inherent in the series. If a judge has an own position that is outside the scale implied by the series, he does not adopt that scale. Instead, he adopts one that is defined by his own position serving as an end anchor at the aborted end of the continuum. Consequently, his scale is broader than that implied by the series, and his judgments accordingly reflect grosser discriminations than



those of his fellow judges who accept the series-defined scale. The negative relationship between judges' attitudes and item values reported by Hovland and Sherif (1952) is interpreted as a result of this difference in the scales adopted by in-range and out-of-range judges.

In his 1929 monograph Thurstone (Thurstone & Chave, 1929) staked the validity of the equal-appearing intervals model on his assumption that scale values are independent of judges' attitudes. On the basis of their evidence against this assumption, Hovland and Sherif (1952) argued that the validity of the model is in doubt. However, the present support for the variable series model suggests that judges' attitudes do not appreciably influence item values unless the item range is narrower than the range of judges' attitudes. From this perspective, it can be argued that Hovland and Sherif actually found only that Hinckley's (1932) item series was not valid for their judges. In theory it is possible to construct an item series which covers the positions of all members of any population of judges. If, in practice, a scale constructor cannot compile such a series, he appears to have the option of specifying a particular population of judges for whom the series can be expected to cover all own positions. This option would probably be satisfactory for most uses of attitude scales.

#### SUMMARY

An experiment was performed to test four alternative interpretations of the influence of judges' attitudes on equal-appearing intervals scale values. The experimental design involved manipulation of the item series as well as judges' attitudes. Two of the theoretical models, perceptual vigilance and assimilation-contrast, can account for the obtained main effect due to judges' attitudes. Another of the models, adaptation level, can account for that effect, as well as the observed effect due to item series. Only one of the models, variable series, can account for both main effects and the significant interaction observed between judges' attitudes and item series.

Although the data generally support variable series, some of the obtained results do not. Notable, in this regard, is the finding of

a positive relationship between own positions and scale values for extremely pro and moderately pro items.

On the basis of the general support for the variable series model, it was argued that valid equal-appearing intervals scales can be constructed only when the own positions of all judges are adequately covered by the item series. It was suggested that scale constructors specify the population of judges for whom this condition is met.

#### REFERENCES

- EDWARDS, A. L. A critique of "neutral" items in attitude scales constructed by the method of equal appearing intervals. *Psychol. Rev.*, 1946, 53, 159-169.
- FEHRER, ELIZABETH. Shifts in scale values of attitude statements as a function of the composition of the scale. *J. exp. Psychol.*, 1952, 44, 179-188.
- FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.*, 1937, 32, 675-701.
- FRIEDMAN, M. A comparison of alternative tests of significance for the problem of  $m$  rankings. *Ann. math. Statist.*, 1940, 11, 86-92.
- HINCKLEY, E. D. The influence of individual opinion on construction of an attitude scale. *J. soc. Psychol.*, 1932, 3, 283-296.
- HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, 55, 244-252.
- HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, 47, 822-832.
- HUNT, W. A., & VOLKMANN, J. The anchoring of an affective scale. *Amer. J. Psychol.*, 1937, 49, 88-92.
- LIKERT, R. A technique for the measurement of attitudes. *Arch. Psychol., N. Y.*, 1932, No. 140.
- MANIS, M. The interpretation of opinion statements as a function of recipient attitude. *J. abnorm. soc. Psychol.*, 1960, 60, 340-344.
- MICHELIS, W. C., & HELSON, H. A reformulation of the Fechner law in terms of adaptation level applied to rating scale data. *Amer. J. Psychol.*, 1949, 62, 355-368.
- SEGALL, M. H. The effect of attitude and experience on judgments of controversial statements. *J. abnorm. soc. Psychol.*, 1959, 58, 61-68.
- THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude*. Chicago: Univer. Chicago Press, 1929.
- VOLKMANN, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951. Pp. 273-294.
- WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid, 1949.

(Received January 3, 1961)

## THEORY AND EXPERIMENT ON THE MEASUREMENT OF APPROACH-AVOIDANCE CONFLICT<sup>1</sup>

SEYMOUR EPSTEIN AND WALTER D. FENZ

*University of Massachusetts*

A number of years ago, Luria (1932) introduced a technique for the measurement of conflict and reaction to stress. Among other procedures, he verified his approach by testing students before an unusually critical examination and by testing prisoners facing the death penalty and other serious sentences. The present study follows in the tradition of Luria's work in that an acute real life event is investigated, an attempt is made to measure affect selectively related to the event rather than the general affective state alone, and a specially devised word association test is used. The approach differs from Luria's in some of the measures investigated, in the substitution of a stimulus dimension for a dichotomous classification of the stimulus, and in the overall theoretical orientation. The critical event is a parachute jump for an inexperienced parachutist.

There is little question but that the first few jumps for a novice parachutist represent an acute approach-avoidance conflict between need for adventure, excitement, or prestige on the one hand, and fear of death and injury on the other. Expressions such as, "It was the most frightening experience of my life," or "For a while I thought I would not be able to go through with it," are commonly heard. Parachuting would seem to be comparable in momentary intensity to the events studied by Luria (1932), and clearly fulfills the conditions of an approach-avoidance conflict. In addition, parachuting has a unique advantage in that, unlike a murder or a critical examina-

tion, it can be repeated within short intervals, thereby allowing for the use of subjects as their own controls and for the investigation of experience and mastery. Finally, the frequency and timing of a parachute jump can be controlled according to the experimental plans of the investigator. In the present study, novice sport-parachutists were tested on the day of an anticipated jump and at a time 2 weeks from a jump. The general influence of parachuting and of time of testing upon responses to a stimulus dimension are investigated in relation to a theoretical model.

### *A Model for the Measurement of Conflict with Projective Techniques*

The model required is one which represents approach as stronger at the goal than avoidance, since subjects by their own choice advance to the goal. Such a model adapted from Miller (1948), is presented in Figure 1. It is assumed that there is an approach drive and an avoidance drive associated with parachute jumping, and that the approach drive can be represented by a less steep gradient as a function of an increasing dimension of stimulus relevance than the avoidance drive. It is assumed that drives have activating and directing properties which can be separately represented. It is assumed that the magnitude of drive produced increment in approach response can be represented by the algebraic difference of the approach and avoidance drives (see Figure 2). It is assumed that approach can be measured by goal relevant verbal responses and avoidance by the failure to produce such responses when they are normally elicited by the stimulus. It should be noted that the model permits the direct prediction of drive produced *increment* in approach response only and not overall drive relevance of the response, which is predominantly a function of the stimulus.

In regard to activation, it is assumed that

<sup>1</sup> This study is part of a project on the measurement of drive and conflict which is being supported by Grant M-1293 from the National Institute of Mental Health, United States Public Health Service. The study was also partially supported by a "teacher's research grant" from the University of Massachusetts.

Epstein is responsible for the theoretical portions of the paper. Fenz, who himself is a parachutist, conducted the study under the supervision of Epstein.



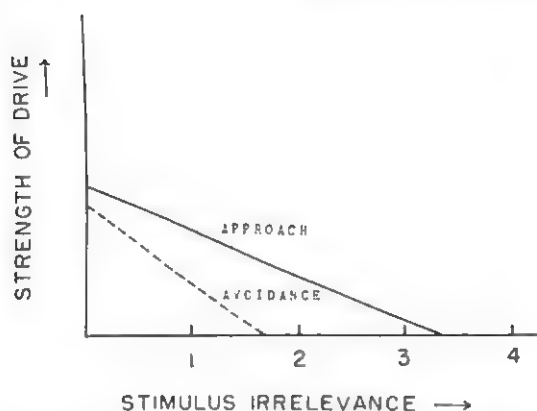


FIG. 1. Strength of approach and avoidance drive as a function of a stimulus dimension.

the magnitude of conflict produced activation can be represented by the sum of the magnitudes of the approach and avoidance drives disregarding algebraic sign (see Figure 3). It is assumed that activation can be measured by the GSR (Woodworth & Schlosberg, 1954), or by some other measure of autonomic activity, or a measure of somatic tension.

One advantage in having the gradients represent drives rather than responses or response tendencies, as in Miller's (1948, 1951) model, is that it immediately suggests the gradients be added if activation is to be considered, and allows for the prediction of two types of event from one basic model. A further reason for assuming a central concept of drive with activating and directive components is that naturalistic observation reveals that a person or animal brought into increasingly close contact (as measured by distance, goal relevant cues, or perceived availability)

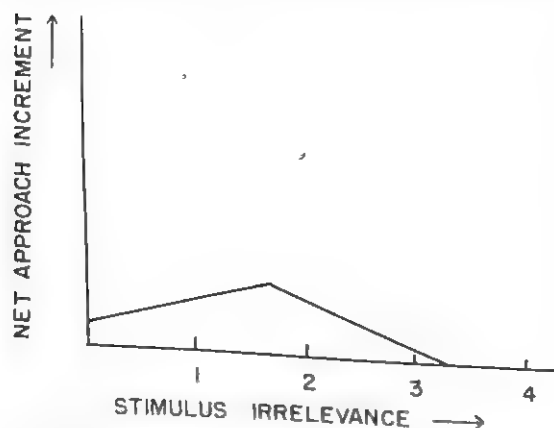


FIG. 2. Net approach increment as a function of a stimulus dimension.

to a goal object exhibits increasing activation as well as stronger approach (or avoidance) responses. It is the unusual child who does not become perceptibly more excited (activated) as a wished-for event (e.g., Christmas morning) approaches, in addition to making approach responses (looking for and questioning about presents). Finally, not only is a central concept of drive helpful in referring to simultaneous changes in activation and goal direction brought about by the presence of a goal object, but, more important, the two aspects of drive are often reducible to each other. In the absence of experimentation it will be necessary once more to turn to observation and example.

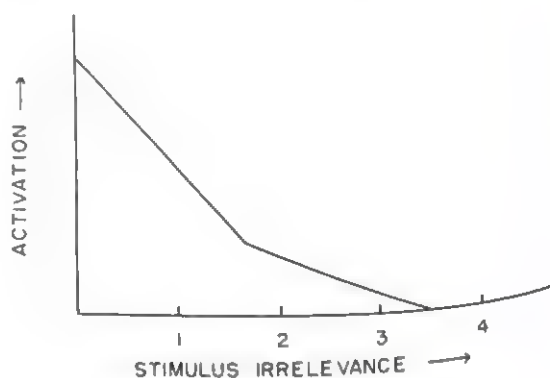


FIG. 3. Activation as a function of a stimulus dimension.

A dog belonging to one of the authors provides an excellent illustration. The animal makes intense approach responses in the form of running and leaping toward a ball, and clearly exhibits a gradient of approach by reacting more vigorously when the ball is near than when it is far. However, if he is commanded to sit still, and the ball made to approach the dog, a gradient of activation is observed in place of a gradient of approach. As the ball is brought closer, the dog becomes perceptibly more tense, his muscles appear to bulge through his skin, and a general tremor appears. We have no doubt but that if physiological measures of activation were recorded a steep gradient of activation would be indicated. Finally, when the ball is almost touching the animal's nose, the tension becomes unbearable and the dog barks vigorously, snaps at the ball, or entirely breaks the inhibition and dashes about the room.

While it is true that Miller allows for gradients of activation by relating avoidance to learned fear, his approach is different from our own in that we assume that activation is

a characteristic of all drives. We hypothesize that a gradient of activation can be demonstrated by using a barrier or physical restraint to prevent approach, but that gradients resulting from inhibition of approach responses are steeper, since they are determined by activation from two drives.

In adapting Miller's model to projective techniques, it is necessary to substitute verbal expression for approach and verbal inhibition for avoidance.<sup>2</sup> This raises the question of whether the principles of conflict that have been worked out by Miller and his colleagues in simple situations can, in fact, be translated into this different and more complex situation. The most critical assumption in the conflict and displacement models is that the gradient of avoidance is steeper than the gradient of approach. Miller (1948, 1951) indicates that this is not always the case but is contingent upon approach being more dependent than avoidance upon a physiological state. He indicates that the conditions are satisfied when approach is based upon hunger, and avoidance upon learned fear. If this is true, it would follow that inhibitory reactions should produce steeper gradients than the drives they inhibit, as the former are contingent upon the latter, and therefore more removed from any primary drive.

Assuming that the gradient of inhibition is steeper than the gradient of expression, the following hypotheses are indicated for the measurement of conflict with stimulus dimensions built into projective techniques:

<sup>2</sup> It has come to our attention, after writing this paper, that Miller (1959), in his more recent work, has also related his model to projective techniques. However, he neither recommends the construction of stimulus dimensions nor the use of measures of activation. He believes that a projective response can be viewed as a displaced response, but that the relationship is complicated by additional considerations. (For an earlier discussion of the construction of stimulus dimensions in projective techniques to measure drive and conflict, see Epstein & Smith, 1956.) We agree with Miller on the complexity of the projective response, and, as will be seen later, find that predicting net approach increment as a function of conflict is theoretically complex in itself, apart from the consideration of other factors. Nevertheless, we believe that the difficulties may not be insurmountable, and even if they are for net approach increment, they will not be for measures of activation.

1. Conflict is indicated by a sharp rise in activation as a function of increasing stimulus relevance.

2. Conflict is indicated by a relative increase in strength of approach responses to stimuli of low relevance and a relative decrease in strength of approach to stimuli of high relevance.

3. Conflict, when of sufficient magnitude, is indicated by a decrease in adequacy of performance as a function of increasing stimulus relevance. This last hypothesis follows from the consideration that high levels of activation are disruptive and that the directive effects of strong expressive and inhibitory drives result in inappropriate overemphasis upon certain stimuli and avoidance of others. The qualification that the conflict must be "of sufficient magnitude" is added to take into consideration that the curve of performance as a function of activation is probably inverted U shaped (Malmo, 1959).

#### *Model Applied to the Present Study*

In the present study, parachutists were tested on the day of a jump and at a time 2 weeks removed from the jump. The model must, therefore, take into account time to a critical event. As a first step, models corresponding to Figures 1, 2, and 3, but with a time dimension substituted for a stimulus dimension, can be considered. If anticipated distance to an event is taken as a measure of distance from the event, the proposed model corresponds to Miller's model for conflict. As the study simultaneously investigates the effects of two dimensions, it is necessary to consider a three-dimensional model. Such a model, which is analogous to combining Miller's models for conflict and displacement, is illustrated in Figure 4.<sup>3</sup> The curve for activation as a function of the stimulus dimension is taken from Figure 3 and reproduced on the z axis. A corresponding curve for activation as a function of a time dimension is represented on the x axis. Four points are selected along the time dimension, and the relationship of level of activation to the stim-

<sup>3</sup> See Murray and Berkun (1955) for an ingenious study in which a three-dimensional model of conflict and displacement is used to predict performance of rats in a runway with displacement alleys.



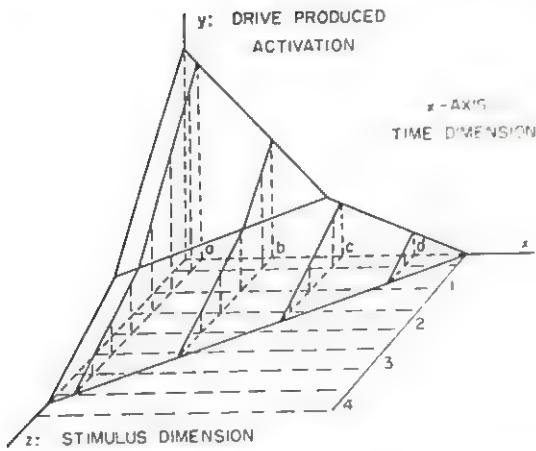


FIG. 4. Activation as a function of a stimulus and a time dimension.

ulus dimension for these points in time is indicated along the surface of the figure. The four resultant curves are extracted and represented to scale in Figure 5, which presents a family of curves for different time intervals to the jump. It is apparent that the gradients become higher and steeper as time to the jump increases.<sup>4</sup>

From Figure 5 and assumptions already made, the following predictions are indicated for the present experiment:

1. Parachutists at both times of testing will produce positive gradients as a function of an increasing stimulus dimension on the following measures: GSR, reaction time, and appropriateness of response. The prediction for reaction time follows from the consideration that for the word association test, response latency is an index of adequacy of response.

2. Parachutists tested on the day of an anticipated jump will produce higher and steeper gradients on each of the above measures than parachutists tested 2 weeks from a jump.

In order to predict the increment in approach response as a simultaneous function of a stimulus and time dimension, it is again necessary to consider a three-dimensional model. The curve for net approach increment

from Figure 2 is reproduced on the  $x$  and  $z$  axes of Figure 6, with the  $x$  axis representing the time dimension and the  $z$  axis the stimulus dimension. The solid figure produced differs from the one for activation in that instead of a change in steepness there is a change in direction of slope. The surface can be described as a triangle leaning into a corner with the apex bent down to meet the intersection of the vertical walls at a point below the point of inflection. Four planes parallel to the  $yz$  plane represent the same four time intervals as in the figure for activation (see Figure 4). The intersection of these planes with the surface of the figure represents the relationship of net approach increment to the stimulus dimension for the time interval in question. The curves are extracted from Figure 6 and presented as a family of curves in Figure 7. Referring to Figures 6 and 7, it can be observed that the change in direction of the slope for both the time dimension and the stimulus dimension makes the prediction of net approach increment much more complex than was the case for activation. Depending upon the point along the time dimension selected, an increase along the stimulus dimension can result in an increase or a decrease in net approach increment. On either side of the point of inflection the same values occur, so that a lack of precision in locating the point of inflection is more critical than was the case for activation. Despite these difficulties, which prevent more specific predictions, the following would seem to be indi-

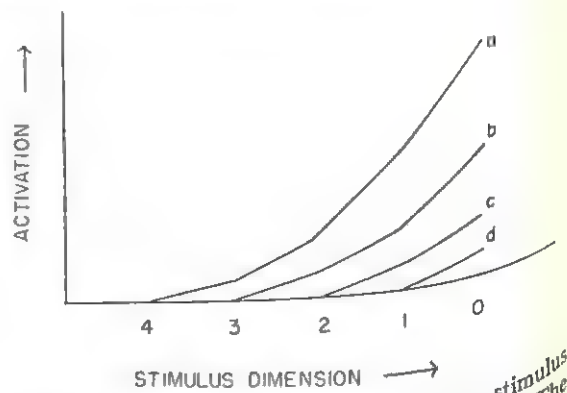


FIG. 5. Activation as a function of a stimulus dimension with time treated as a parameter. (The letters correspond to the letters in Figure 4, with a representing the shortest, and d the longest time interval.)

<sup>4</sup> It should be noted that the change in steepness of the slopes is a consequence of adding the two gradients. If either gradient alone were represented, a series of parallel curves differing only in height would be obtained. This may be observed in Curves c and d, which do not contain a component of the avoidance drive.

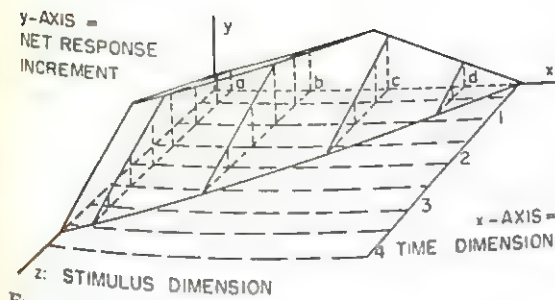


FIG. 6. Net approach increment as a function of a stimulus and a time dimension.

cated for the present study: (a) Parachutists at both times of testing will produce more parachute relevant responses than nonparachutists to words at the lower end of the stimulus dimension. (b) Parachutists on the day of an anticipated jump will produce more parachute relevant responses to stimuli of low parachute relevance and less parachute relevant responses to stimuli of high parachute relevance than parachutists tested 2 weeks from a jump, i.e., there is a significant interaction between time to a jump and stimulus relevance.

## METHOD

### Subjects

Experimental subjects were 16 novice parachutists from the University of Massachusetts Sport Parachuting Club. All had made at least one and not more than three parachute jumps at the time of testing. The subjects were recruited at the regular meeting of the parachute club and were asked to take part in the experiment for the payment of the costs of one parachute jump. The experimental group was matched with a comparable group of control subjects selected from students in an elementary course in psychology at the University of Massachusetts who volunteered to take part in experiments. Three additional subjects, who had considerable experience with parachuting, were also tested.

### Materials

Parallel forms of word association tests were constructed containing words with four levels of relevance to parachuting. The words were selected from a larger group rated on a four-point scale of parachute relevance by three experienced parachutists. Only words on which full agreement was reached were retained. The words were presented in increasing order of relevance to prevent highly relevant words from influencing responses to words of lower relevance. Three anxiety words were added to each list mainly out of curiosity, and were placed at the end of the list in order not to interfere with the dimension. One or more neutral words was

presented between parachute relevant words and between anxiety words, and 10 neutral words were presented at the beginning of the list to reduce practice effects. Following are the word lists, with level of parachute relevance designated by figures, except for neutral words which were treated as the lowest level of relevance, and with anxiety words designated by the letter a:

**Word List A:** SALT, QUIET, RED, JOY, FOOTBALL, CITIZEN, MUSIC, CHAIR, SWISS CHEESE, HUNGRY, SKY<sub>1</sub>, RHINOCEROS, LAND<sub>1</sub>, BOOK, CATTLE, MOUNTAIN<sub>1</sub>, MOON, WHITE, FALL<sub>2</sub>, LOUD, DOG, FLOAT<sub>2</sub>, SCISSORS, ALTITUDE<sub>2</sub>, LION, SWIFT, BAIL-OUT<sub>2</sub>, STOVE, RADIATOR, CANOPY<sub>2</sub>, TAXI, GREEN, TARGET<sub>2</sub>, SOUR, FEAR<sub>2</sub>, BEAUTIFUL, BLACKOUT<sub>2</sub>, MAN, INJURY<sub>2</sub>, CARPET, HAPPY.

**Word List B:** MONKEY, MOON, RED, ELEPHANT, KEY, KING, COMFORT, SLEEP, SCREWDRIER, CHICKEN, WIND, GREEN, SAND<sub>1</sub>, PAPER, PENCIL, HILL<sub>1</sub>, BLACK, STAR, DESCEND<sub>2</sub>, HEAVY, FOOTBALL, OPENED<sub>2</sub>, HAMMER, HARNESS<sub>2</sub>, CAT, HARD, JUMP<sub>2</sub>, LAMP, EAT, PARACHUTES, TOWN, PINK, RIPCORD<sub>2</sub>, BITTER, ANXIETY<sub>2</sub>, SOFT, KILLED<sub>2</sub>, BOY, HURT<sub>2</sub>, TABLE, FUN.

Words were presented by a tape recorder at 30-second intervals after having been screened by four judges for clarity of pronunciation. Subjects were instructed to respond as quickly as possible with the first word that occurred to them. Response words were recorded by the experimenter. Recordings of skin resistance were obtained from a direct current Hunter GSR apparatus, operating on the principle of a Wheatstone bridge. The apparatus was adapted for finger electrodes, and resistance was registered by an automatic recording pen. Time from end of stimulus to beginning of response was recorded by a pen-marker controlled by the experimenter.

### Procedure

Eight experimental subjects were tested on the day of a jump, and again 2 weeks after the jump; eight others were tested 2 weeks before a scheduled jump, and again on the day of the jump. Subjects were required to refrain from jumping for 2 weeks preceding the "before jump" testing and for 2

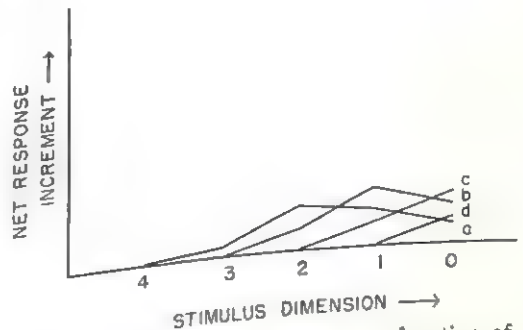


FIG. 7. Net approach increment as a function of a stimulus dimension with time treated as a parameter. (The letters correspond to the shortest, and d the longest with a representing the shortest, and d the longest time interval.)



weeks following the "after jump" testing, and to make no additional jumps between the testing sessions. Testing on the day of a jump took place 1-3 hours preceding the jump. Sixteen control subjects were tested twice with a 2-week interval between testing. In all cases, List A was presented in the first session and List B in the second session.

Upon entering the experimental room and being seated facing the examiner, a brief explanation was given about the word association test, during which the subject was told that the purpose of the test was to measure emotions related to parachuting. Electrodes were then attached to the right hand index and middle fingers. After about 3 minutes, to permit the electrodes to polarize and the subject to become accustomed to them, the tape recorded list of words was introduced with the comment:

Now we will begin a test of speed of association to words. After you have heard a word—and not before you have heard the whole word—say the first word that occurs to you as quickly as possible. If you are not sure of a word, respond to what you think it was. Please do not make comments or ask questions between words, but save them for the end of the test. Any Question?

Where distributions permitted, the data were evaluated by analyses of variance. In one analysis, experimental subjects were used as their own controls, and nonparachutists were omitted. This permitted an orthogonal arrangement of experimental condition and order. In a second analysis, control subjects were compared on first and second testing.

## RESULTS

### *Galvanic Skin Response*

*Response to the stimulus dimension.* Galvanic skin response was measured by change in conductance immediately following presentation of a stimulus word. In order to make the results independent of selective perception, which in itself was found to be a significant variable, GSR was analyzed only for words which had not received a score for misperception. In addition, GSR scores were eliminated for words in the stimulus dimension when it was clear that the subject interpreted a word in a sense other than the one on which it had been dimensionalized, e.g., the response "winter" to FALL. Finally, GSRs to neutral words which elicited parachute relevant responses were discarded, as their neutral status had become questionable. By eliminating the complicating effects of perception, we could reasonably assume a meaningful stimulus dimension for each of the experimental subjects, instead of a dimension

of the probability that the word is related to parachuting. Luria (1932) follows the latter procedure when he categorizes his stimuli as critical, doubtful, or indifferent. In order to correct for the resultant unequal number of responses in a category scores were represented by average GSR per category. Actually, very few words were discarded, never more than one word in any of the categories other than neutral, for which two was the maximum. The results would be essentially the same, but more difficult to interpret, had all words in a category been used. In representing the neutral words, the first 10 "practice" words were not counted.

Sixteen out of 16 experimental subjects produced a gradient of increasing GSR as a function of increasing stimulus relevance. One control subject also produced a gradient, but on questioning it turned out that he had been a paratrooper in the Army. His data were replaced by that of another control subject. For every experimental subject the gradient on the day of a jump was higher and steeper than on a control day. Although statistical analysis might appear to be superfluous, analyses of variance were carried out to investigate the effects of sequence, order, individual differences, and to allow for a comparison with the analyses of reaction time and of anxiety words. In addition, since the findings on the gradients were unequivocal, an opportunity was afforded for testing the relative validity of three commonly used indices of GSR.<sup>5</sup>

<sup>5</sup> The three measures were change in resistance as measured in ohms, change in conductance as measured in mhos, and change conductance as measured in log mhos. Results obtained from each measure were plotted for each experimental subject. On all three measures 16 of the 16 subjects showed a gradient of response corresponding to the dimension of stimulus relevance, and for each subject the gradient on the day of the jump was steeper than on the day remote from the jump.

Analysis of variance for each separate transformation yielded parallel results. A comparison of the three measures indicated that *F* values for log conductance change were less pronounced than for resistance change and conductance change. This discrepancy may be attributed to the fact that the logarithmic measure reduced the distance between highly divergent scores. Logarithmic transformation is commonly used to reduce the relationship between conductance change and prestimulus level. Inspection

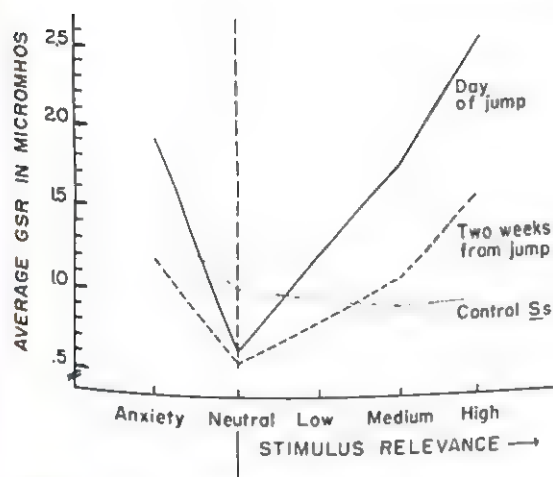


FIG. 8. GSR of parachutists and controls to a stimulus dimension and to anxiety words.

It is apparent that the control group produces no gradient (see Figure 8). The mean GSR in micromhos is 1.04 for neutral words and .95 for words of high relevance. The parachutists produce a gradient on the control day and a higher and steeper gradient on the day of a jump, all of which is in accordance with prediction.<sup>6</sup> The mean GSR

of the data indicated a positive relationship between prestimulus conductance level and magnitude of the deflection. In order to determine whether the results on GSR for the transformed measures of conductance could be attributed to differences in prestimulus level, the same analyses of variance were run on average prestimulus level as on GSR. The analysis of variance for prestimulus levels of experimental subjects on the day of the jump against prestimulus levels on the day remote from the jump did not yield a difference approaching significance. When experimental subjects under both conditions were compared to control subjects the analysis of prestimulus conductance measures also failed to yield significant values. The results for conductance change therefore could not be attributed to absolute conductance levels. That findings which were conclusively present by inspection were measured with least reliability by log conductance units throws some doubt on the advisability of this transformation for GSR data.

The differential steepness of the gradients for the three groups cannot be attributed to the unequal heights of the gradients. The curves for the parachutists intersect the curve for the controls, and each other within the same range on the ordinate. If anything, the difference in the heights of the gradients can be attributed to the difference in slopes, rather than the reverse, as the parachutists on the day of a jump and a control day do not differ significantly on neutral words.

in micromhos for neutral and high relevant words, respectively, is .56 and 1.57 on a control day, and .65 and 2.59 on the day of a jump.

Only the analysis of variance for parachutists serving as their own controls is presented, as no source of variance approached significance for the analysis of control subjects. In Table 1 it can be seen that there are significant differences for experimental condition, for stimulus dimension, and for the interaction of experimental condition and stimulus dimension. Significance is also indicated for the main effect of individual differences and for the interaction of individual differences with experimental condition. Apparently, not only do individuals differ reliably in the overall magnitude of their GSRs, but they also differ in the effect an approaching critical event has upon their GSRs. Perhaps this indicates an interesting personality variable that should be followed up in further work.

An additional finding of interest is that the parachutists have a smaller mean GSR than the control subjects to neutral words on both the day of a jump and the control day (see Figure 8). When tested by a separate analysis of variance for neutral words only, the difference is found to be significant at the .01 level.

TABLE 1  
ANALYSIS OF VARIANCE OF GSR TO A STIMULUS  
DIMENSION FOR PARACHUTISTS ON THE DAY  
OF A JUMP AND A CONTROL DAY

Source of variance	df	MS	F
Between-subjects	15	129.41	1.15
Sequence (Q)	1	111.70 <sup>a</sup>	11.67***
Subjects	112		
Within-subjects	1	795.51	17.07**
Group (G)	1	46.92	1.00
Q × G	14	46.58 <sup>a</sup>	4.87***
Subjects × G	3	1309.91	95.12***
Dimension (D)	3	11.01	.79
Q × D	42	13.77 <sup>a</sup>	1.44
Subjects × D	3	143.71	15.02***
G × D	3	5.43	.56
Q × G × D	42	9.57 <sup>a</sup>	
Subjects × G × D	127		
Total			

<sup>a</sup> Error term for mean squares above it up to the next error term. The Subjects × G × D interaction was used to evaluate all other error terms.

\*\* Significant at .01 level.

\*\*\* Significant at .001 level.



TABLE 2

ANALYSIS OF VARIANCE OF GSR TO NEUTRAL AND ANXIETY WORDS FOR PARACHUTISTS ON THE DAY OF A JUMP AND A CONTROL DAY

Source of variance	df	MS	F
Between-subjects	15		
Sequence (Q)	1	133.58	2.26
Subjects	14	59.00 <sup>a</sup>	6.88***
Within-subjects	48		
Group (G)	1	267.62	4.46
Q × G	1	50.70	.84
Subjects × G	14	59.94 <sup>a</sup>	6.97***
Category (C)	1	1483.10	29.17***
Q × C	1	72.58	1.42
Subjects × C	14	50.85 <sup>a</sup>	5.91**
G × C	1	165.53	19.27***
Q × G × C	1	9.61	1.12
Subjects × G × C	14	8.59 <sup>a</sup>	
Total	63		

<sup>a</sup> Error term for mean squares above it up to the next error term. The Subjects × G × C interaction was used to evaluate all other error terms.

\*\* Significant at .01 level.

\*\*\* Significant at .001 level.

**Reaction to anxiety words.** A comparison of mean GSR to neutral and anxiety words (see Figure 8) indicates that both experimental and control subjects react more strongly to anxiety words than to neutral words. This difference is greatest for experimental subjects on the day of a jump, with a mean difference in micromhos of 1.28, next for experimental subjects on the day remote from a jump, with a mean difference of .64, and least for control subjects, with a mean difference of .41. As with neutral words the parachute group on a control day has a lower GSR to anxiety words than the control group. The analysis of the parachutists as their own controls is presented in Table 2, where it can be seen that not only do parachutists react significantly differently to anxiety and neutral words (C, .001 level) and that their differential reaction varies as a function of proximity to the jump (G × C, .001 level), but there are also significant individual differences in GSR across conditions (subjects), in change in GSR from control day to jump-day (Subjects × G), and in difference in GSR between neutral and anxiety words (Subjects × C). In the analysis of control subjects by themselves the difference between anxiety and neutral words is significant at the .05 level.

It is noteworthy that the parachutists on both the day of a jump and a control day

produce larger GSRs to high relevant words than to anxiety words, which is consistent with the assumption that the gradients represent activation rather than fear alone.

### Reaction Time

**Response to the stimulus dimension.** Reaction time produced similar results to GSR. In Figure 9 it can be seen that at both times of testing the parachute group produces a positive gradient of reaction time as a function of the stimulus dimension, and that the gradient is higher and steeper on the day of a jump than on a control day. On the control day, mean reaction time is 1.41" for neutral words and 2.32" for words of high parachute relevance. On the day of a jump the corresponding values are 1.45" and 3.40". The control subjects produce no gradient. Analysis of variance for the parachutists reveals that the difference associated with experimental condition (G), with the stimulus dimension (D), and with the interaction of experimental condition and stimulus dimension (G × D) are all significant (see Table 3). No source of variance is significant in the analysis of variance for the control group.

Of additional interest, and parallel to the findings in GSR, is the indication in Figure 9 that mean reaction time of parachutists to neutral words on a control day ( $M = 1.41''$ ) is less than for control subjects ( $M = 1.69''$ ), although in this case the analysis of variance of neutral words alone fails to reach significance.

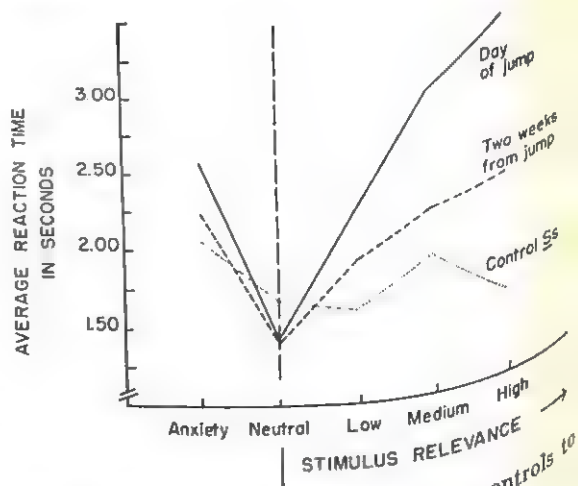


FIG. 9. Reaction time of parachutists and controls to a stimulus dimension and to anxiety words.

TABLE 3

ANALYSIS OF VARIANCE OF REACTION TIME TO A STIMULUS DIMENSION FOR PARACHUTISTS ON THE DAY OF A JUMP AND A CONTROL DAY

Source of variance	df	MS	F
Between-subjects	14		
Sequence (Q)	1	1.34	.42
Subjects	13	3.12	8.66***
Within-subjects	112		
Group (G)	1	10.88	15.32**
Q × G	1	2.13	3.00
Subjects × G	13	.71 <sup>a</sup>	1.97*
Dimension (D)	3	12.62	22.53***
Q × D	3	.18	.32
Subjects × D	39	.56 <sup>a</sup>	1.55
G × D	3	1.62	4.50**
Q × G × D	3	.48	1.33
Subjects × G × D	39	.36 <sup>a</sup>	
Total	119		

<sup>a</sup> Error term for mean squares above it up to the next error term. The Subjects × G × D interaction was used to evaluate all other error terms.  
\* Significant at .05 level.  
\*\* Significant at .01 level.  
\*\*\* Significant at .001 level.

*Reaction to anxiety words.* A comparison of mean reaction time to neutral and anxiety words reveals that it took all groups longer to respond to anxiety than to neutral words (see Figure 9). The difference is greatest for parachutists on the day of a jump (1.12"), next for parachutists on a control day (0.84"), and least for control subjects (0.37"). The results are similar to those obtained for GSR (compare Figures 8 and 9). The findings are confirmed by analysis of variance. The analysis of parachutists on the day of a jump and a control day yields significance at the .01 level for experimental condition and for stimulus category, and significance at the .05 level for the interaction of experimental condition with stimulus category (see Table 4). The analysis of control subjects by themselves yields significance at the .05 level for the difference between neutral and anxiety words.

### Failures in Perception

*Perception of stimulus dimension and anxiety words.* Inspection of the data made it apparent that a considerable number of misperceptions and very few other obviously poor responses occurred. Failures in perception were scored when the subject reported that he had not heard a word, or when a response was given that was more appropriate

TABLE 4

ANALYSIS OF VARIANCE OF REACTION TIME TO NEUTRAL AND ANXIETY WORDS FOR PARACHUTISTS ON THE DAY OF A JUMP AND A CONTROL DAY

Source of variance	df	MS	F
Between-subjects	14		
Sequence (Q)	1	.14	.07
Subjects	13	2.07 <sup>a</sup>	2.46
Within-subjects	45		
Group (G)	1	13.20	9.49**
Q × G	1	.73	.86
Subjects × G	13	1.39 <sup>a</sup>	1.65
Category (C)	1	8.48	9.31**
Q × C	1	.64	.76
Subjects × C	13	.91 <sup>a</sup>	1.08
G × C	1	4.08	4.85*
Q × G × C	1	.31	.36
Subjects × G × C	13	.84 <sup>a</sup>	
Total	59		

<sup>a</sup> Error term for mean squares above it up to the next error term. The Subjects × G × C interaction was used to evaluate all other error terms.  
\* Significant at .05 level.  
\*\* Significant at .01 level.

to a word that sounded like the stimulus word than to the stimulus itself, such as the response "person" to sky, where later questioning confirmed that the subject had heard sky as "guy." Almost all inferred misperceptions were ones that occurred several times, and had been confirmed at least once by later questioning.

Figure 10 presents number of words missed over total words in a category, expressed as a percentage. Low and moderate parachute relevant words were combined to increase

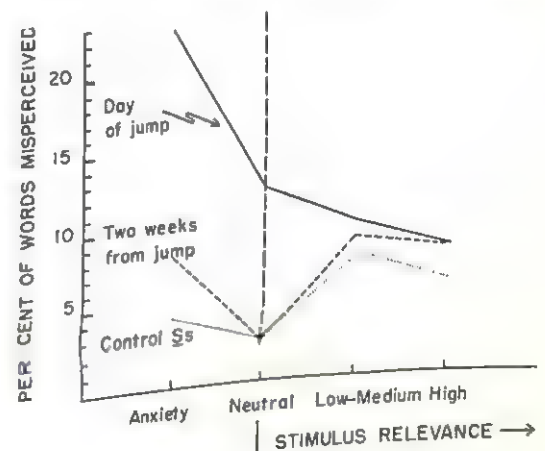


FIG. 10. Percentage of misperceptions by parachutists and control subjects of words on a stimulus dimension and anxiety words.



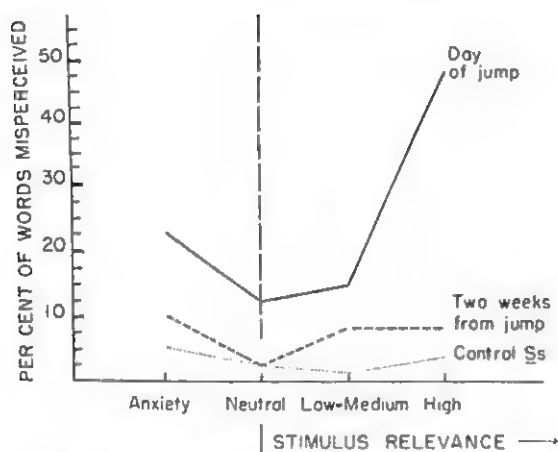


FIG. 11. Percentage of misperceptions of parachute words and anxiety words following words on a stimulus dimension and anxiety words.

reliability. Neutral words were restricted to the first 10 practice words because inspection (later verified) suggested that misperception of neutral words was influenced by the preceding parachute relevant word. It is evident in Figure 10 that parachutists on the day of a jump manifest a general perceptual deficit, as indicated by misperception of neutral words. There appears to be a specific deficit for anxiety words, and a relative sensitization<sup>7</sup> to parachute relevant words in relation to degree of relevance.

The high number of cases with zero misperceptions produced highly skewed distributions and accordingly chi square analysis of frequency data was used. A comparison between parachutists on the day of a jump and a control day revealed that there were 10 parachutists who produced an unequal number of misperceptions of the first 10 neutral words on the day of a jump and a control day and 6 who produced an equal number of misperceptions. All of the former produced more misperceptions on the day of a jump (significant at the .01 level). Of 11 parachutists who produced an unequal difference between misperception of neutral and high parachute relevant words on the day of a jump and a control day, 9 made relatively fewer misperceptions of parachute words on the day of a jump, which is significant at the

<sup>7</sup> Perceptual deficit and sensitization are used here descriptively only, and are not meant to imply a reason for the phenomenon.

.05 level. In addition, 12 out of 16 control subjects as compared to 6 out of 16 parachutists on the day of a jump produced a higher number of misperceptions of high parachute than neutral words, which is significant at the .05 level. Of 10 parachutists who misperceived an unequal number of anxiety words on the day of a jump and a control day, 9 made more misperceptions on the day of a jump, which is significant at the .02 level. Moreover, of 14 parachutists who produced an unequal difference in misperception of neutral and anxiety words on the day of a jump and a control day, 11 misperceived relatively more anxiety words on the day of a jump, which is significant at the .05 level. When comparison was with control subjects, 2 out of 16 controls and 11 out of 16 parachutists on the day of a jump misperceived more anxiety than neutral words, which is significant at the .01 level. Thus, on the basis of statistical analysis it may be concluded that parachutists on the day of a jump manifest a general perceptual deficit, as indicated in their misperception of neutral words, a specific deficit for anxiety words, and relative perceptual sensitization for parachute words.<sup>8</sup>

*Perception of neutral words following relevant words.* It has been noted that although

<sup>8</sup> The findings cannot be explained away by intra- or intralist set effects. For one thing, if there were such set effects, they should influence the control subjects, yet there is no evidence of any perceptual effect for the control subjects. Second, parachutists served as their own controls, with half tested first on the day of a jump, and half tested first on a control day, and a careful analysis indicated that both subgroups behaved similarly. So far as intraserial set effects are concerned, the neutral words were presented before the parachute words, so that influence on the former was not possible. It might be argued that intraserial set effects could account for the sensitization to parachute words and the relative deficit for anxiety words, which appeared last. However, this would influence control subjects as well as parachutists on a control day, and cannot be a factor unless one is willing to postulate differential sets to pick up sets on the day of a jump, which is not too far off from selective perception. To determine whether such was the case, we examined the responses to successive words in the same category. There was no evidence for differential group performance as a function of an emerging set. It may be concluded that the data cannot be accounted for, directly or indirectly, by intra- or interlist set effects.

parachutists on the day of a jump had no difficulty in perceiving parachute related words they tended to misperceive the words that immediately followed. Accordingly, perception of neutral words was investigated as a function of the words preceding them. Figure 11 presents percentage of failures in perception for each of the categories of neutral words. It is evident that parachutists on the day of a jump produce a positive gradient of failures in perception as a function of an increasing stimulus dimension. There is also a tendency for parachutists on the day of a jump to misperceive neutral words following anxiety words.

Of 12 subjects who misperceived an unequal number of neutral words following words of high parachute relevance on the day of a jump and a control day, 10 had more misperceptions on the day of a jump, which is significant at the .05 level. When comparison is between control subjects and parachutists, 4 out of 16 controls and 10 out of 16 parachutists on the day of a jump misperceive at least one neutral word following a word of high parachute relevance, which is again significant at the .05 level. A comparison of misperception of neutral words following words of high relevance, relative to misperception of first 10 neutral words, reveals that of 15 parachutists who produced an unequal difference on the day of a jump and a control day, 12 made relatively more misperceptions of neutral words following parachute words on the day of a jump, which is significant at the .05 level. When the comparison is between parachutists and control subjects, 2 out of 16 controls and 9 out of 16 parachutists on the day of a jump misperceive relatively more neutral words following words of high parachute relevance than words among the first 10 neutral words, which is significant at the .01 level. It may be concluded that on the day of a jump there is a significant increase in the misperception of neutral words following words of high parachute relevance.<sup>9</sup>

<sup>9</sup> The possibility of intra- and interlist set effects producing the results was considered in the same manner as for the words in the dimension. Again, the results were negative. Luria (1932) also reports cognitive disturbance on neutral words following critical words.

Analysis of neutral words following anxiety words revealed that of eight parachutists who misperceived an unequal number of these words on the day of a jump and a control day, seven made more misperceptions on the day of a jump, which is significant at the .05 level. A comparison of misperception of neutral words following anxiety words relative to misperception of first 10 neutral words for parachutists on the day of a jump and either parachutists on a control day or control subjects approached, but did not reach significance ( $p = .05-.10$ ). It may be concluded that there is an unreliable tendency for parachutists on the day of a jump to manifest perceptual deficit for neutral words following anxiety words.

### *Content of Associations*

It was predicted that parachutists on the day of a jump relative to a control day would demonstrate an increase in parachute relevant responses to stimuli of low relevance and a decrease in parachute relevant responses to stimuli of high relevance. Unfortunately, the prediction could not be tested at the high end of the dimension as stimuli of high relevance almost always elicited responses that were relevant to parachuting, and we were unable to devise a reliable scoring system for determining degree of approach within this range. As for the rest of the dimension, there was no evidence that responses to neutral words and to words of low and moderate parachute relevance were functionally dissimilar. Accordingly, responses to all stimuli were combined into one overall score.

Analysis of variance of the responses of parachutists on the day of a jump and a control day revealed a significant difference between the groups at the .01 level. Parachutists on the day of a jump produced an average of 8.3 parachute relevant words as contrasted with 5.8 on a control day. Control subjects produced the least number of parachute relevant words, with a mean of 3.2.

### *Three Records of Experienced Parachutists*

Although the reliability of results based on a sample as small as three is necessarily low, the results were so striking that we decided



they were worth reporting. For both mean GSR and mean reaction time, the gradients for the experienced and inexperienced parachutists were alike on the control day. However, while the inexperienced jumpers produced a monotonic curve on the day of a jump, the experienced jumpers produced a curve similar to it up to an intermediate point on the stimulus dimension, after which it dropped off so sharply that the mean reaction at the upper end of the dimension was no different from what it was at the lower end of the dimension. All 3 experienced jumpers and none of the 16 inexperienced jumpers or 16 controls produced this curve.<sup>10</sup>

## DISCUSSION

### *Measures of Activation*

It had been assumed that GSR is a direct measure of activation and that reaction time in a word association test is a measure of adequacy of performance, and therefore, indirectly of activation. Accordingly it was predicted that parachutists would produce similar gradients for both measures on a control day, and that the gradients would be higher and steeper on the day of a jump. Apart from supporting the model, the high reliability of the findings with these two measures indicates that they are completely adequate for measuring individual differences. Considering that reaction time and particularly GSR are indirect measures of conflict and not normally under conscious control of the individual, they are particularly promising measures of unconscious conflict. The results are striking enough to warrant testing of known clinical groups with specially devised stimulus dimensions used in conjunction with measures of GSR and reaction time. As an example, ulcer patients might be tested with stimulus dimensions of hostility and dependency built into a word association test or a TAT, and schizophrenics might be tested with stimulus dimensions of nurturance and rejection. A study along the latter lines is presently under way. The approach need not be confined to broad

motive states, but specific sources of conflict can be explored as well.

The finding that reaction time produced results almost identical to GSR is consistent with Lanier's (1941) conclusion that the affective change brought about by words representing a conflict influences length of reaction time in the same direction and to about the same extent as GSR. However, it is evident that the relationship between GSR and reaction time is not a direct and simple one. Hathaway (1929) reported a correlation of .60 between GSR and reaction time, while Hunt and Landis (1935) found no relationship. Peterson and Jung (1907) reported that only in certain cases is there a clear relationship. What is probably the case is that GSR and reaction time measure different aspects of conflict. Following the model presented, GSR is a measure of activation, while reaction time, at least in a word association test, is a measure of adequacy of performance, which is a frequent but not necessary concomitant of activation. Assuming that the curve of level of performance as a function of increasing activation is inverted U shaped, the relationship between reaction time and activation would depend in part upon the nature of the task, and in part upon the degree and range of activation sampled. Moreover, in a test where the subject was not required to respond rapidly, reaction time would not be a measure of adequacy of performance. Finally, although increases in activation within the above limitations should increase reaction time, the reverse need not hold and reaction time could be affected by many sources that have no bearing on GSR. To the extent that both measures in a word association test are more highly correlated with conflict than with each other, they should provide useful supplementary measures of conflict.

### *Measures of Approach and Avoidance*

*Perceptual sensitization and defense.* It had been predicted that positive gradients would be produced by measures of misperception and reaction time on the assumption that both were measures of adequacy of response. Although the prediction was substantiated for reaction time, a significant opposite relationship was found for misperception. Yet, the

<sup>10</sup> We have now, as part of another study, confirmed the results with three more experienced parachutists.

basic assumptions from which the prediction for misperceptions was made were supported. i.e., the findings on GSR provided evidence of an increase in activation along the stimulus dimension, and the misperception of neutral words following words on the dimension provided evidence that activation interferes with perception. What had not been taken into account is that perception of words along a dimension is a measure of selective content in much the same manner that content of association is. In the present study approach was greater than avoidance, so that net approach increment could be expected to compensate for interference effects from activation. It is as if the subject, despite a general state of high tension and further increases in tension brought about by thoughts of parachuting, nevertheless forces himself to concentrate on parachuting. The effect is to overcome anxiety induced perceptual deficit for parachute relevant stimuli. The adaptive significance of such a reaction is apparent. It is not possible to explain away the findings by word frequencies, familiarity with words, or inter- or intra-serial set effects. Thus, the evidence would appear to support unequivocally the phenomenon of perceptual sensitization. The findings on anxiety words equally well support selective perceptual deficit, which in light of the experimental controls and negative findings on set effects, is strongly suggestive of perceptual defense, i.e., an anxiety reducing defense mechanism. On the day of a jump novice parachutists are in a state of acute anxiety, with which they must come to grips if it is not to interfere with their performance and endanger their lives (Walk, 1959 unpublished). Thinking about and being oriented toward anxiety producing stimuli, i.e., worrying, would only increase level of anxiety, so that there is a strong basis for an avoidance reaction to such thoughts and perceptions. In order to check the anxiety reducing effects of misperception further, we examined the GSRs to anxiety words that were misperceived. Where an innocuous word was perceived in place of an anxiety word, the GSR was reduced. Thus, the overall evidence suggests that perceptual defense did occur for the anxiety words. Although originally unanticipated, perceptual sensitization and defense

may well provide effective measures of approach and avoidance. Of course, in situations where unconscious conflict was involved, avoidance would be stronger than approach, and net approach tendency would work in the same direction as activation, so that enhanced perceptual defense might be the only phenomenon that would be found.

It is noteworthy that failures in perception were far more common than all other inadequate responses combined. Luria (1932) noted failures in perception for emotionally disturbing words, but these were no more common than were other inadequate responses, such as perservation, repeating a stimulus word, making an incidental comment, or blocking. The larger number of misperceptions in the present study undoubtedly is a result of the tape recorded presentation, despite the screening of the lists for clarity of pronunciation. Tape recordings obviously provide reduced cues because of the absence of lip movements. It is possible that tape recorded lists containing specially selected categories of words would provide an excellent projective technique for general usage, particularly if fidelity or loudness were varied to provide an optimum level of ambiguity or a measure of perceptual threshold.

*Content of association.* The results on content of association corresponded to those on perceptual sensitization, i.e., parachutists on the day of a jump produced an increased number of parachute related responses. Unfortunately, the critical prediction so far as the model is concerned, that conflict would be indicated by an increment in parachute relevant responses at the low end of the parachute dimension and a decrement at the high end, could not be adequately tested because all associations to the high end of the dimension were uniformly high in relevance. To test the hypothesis it will be necessary to use words at the upper end of the stimulus dimension which facilitate avoidance reactions to a greater extent than the ones used in the present study. Nevertheless, the reduced number of misperceptions at the high end of the dimension leads one to suspect that the hypothesis would not be substantiated. Moreover, there was no evidence within the range of the dimension that could be adequately



tested indicating that responses to different levels of the dimension were functionally dissimilar.

The measures of perception and of content of association clearly illustrate the manner in which conflict serves to focus attention in the area of conflict and to produce insensitivity in other areas. The implications for psychopathology are evident, particularly in reference to symptoms involving preoccupation, such as obsessive thinking, and to symptoms of inefficiency and inability to concentrate.

### *Effect of Experience on the Shape of Gradients*

The findings on three experienced parachutists are particularly interesting. A reasonable prediction would have been that experienced jumpers would produce gradients that were lower and less steep than inexperienced jumpers, particularly on the day of a jump, i.e., their gradients would be somewhere between an inexperienced jumper's and a nonparachutists'. However, it was found for both GSR and reaction time that on the day of a jump, experienced jumpers produced similar gradients to inexperienced jumpers up to a midway point along the dimension, after which their gradients reversed direction, so that their GSRs to high relevant parachute words were no different than to neutral words. The resulting GSR and reaction time curves for the experienced jumpers resembled the predicted curve for net approach increment, which was based on the assumption that inhibitory tendencies are in conflict with expressive tendencies. This may indicate that associated with experience and mastery of conflict there is an inhibition of anxiety producing responses. Another possibility, not incompatible with the first, is that the gradient of activation is steeper for cues that are consciously associated with the conflict than for cues that are indirectly conditioned to the conflict, and that with successful experience, the anxiety associated with the labeled cues is extinguished at a faster rate than the anxiety conditioned to the unlabeled cues. Parachutists, in general, are more apt to consciously relate conflict produced by parachuting to high relevant words, such as PARACHUTE and RIPCORDER, than to low relevant

words, such as LAND and HILL. It would then follow that with successful experience conflict produced fear should be extinguished faster to the high relevant words than to the low relevant words. If the argument here is correct, it offers an explanation of why anxiety is often elicited by apparently innocuous or unrecognized stimuli, and why it is therapeutic to label stimuli associated with a conflict. In reference to this latter point, an ex-paratrooper colleague told us of how he had experienced an anxiety attack while standing on the platform of a train, until it occurred to him that the perception of the terrain moving by was similar to his perceptions when preparing to jump from an airplane.

### *Anxiety Reducing Effects of Parachuting*

Parachutists obtained lower GSRs and shorter reaction times to neutral stimuli than did control subjects. Possibly, this indicates a personality characteristic which leads people to choose parachuting as a sport. Short reaction times and small GSRs suggest a person who is relatively impulsive and expressive. A second possibility is that parachuting serves to reduce tensions, and that the focusing of conflict and its momentary mastery in one area temporarily reduces conflict and disturbance in other areas, i.e., parachuting is a way of mastering fear. A few records of parachutists tested immediately after a jump revealed an astonishing absence of GSRs. The father of one of the parachutists reported that his son, who is a chronic stutterer, experienced an alleviation of symptoms on the afternoon following his first jump. Another subject, in a letter to the examiner, volunteered information about his increased frustration tolerance following a parachute jump. Reference is here made to the letter, which provides a description of a first jump:

Well, my first static line jump was an experience that left some definite impressions on me . . . Yes, fear of the unknown was there. I had prepared for this a long time. It was a real challenge to me. I had fears of chickening out, fear that the chute might not open, or that my landing would be hard. I was very anxious to get it over with. The most immediate sequence that is constantly on my mind is the time interval of standing on the wheel and holding on to the struts, being hit by the jumpmaster, and falling backwards till my chute caught.

What happened after that is only slightly in my mind. After landing I was happy, emotionally released, and proud of myself in that I had accomplished something I set out to do. This was a good morale booster and reduced my insecurity . . . While driving back we tried to pick up two girls on the road. They refused, but this did not shatter my ego. I had an "I don't care" attitude. While on the road I did not speed, or at least there was no desire to speed. Another change I noticed was that when I put on the taperecorder to hear some music, I did not want to hear it loud. That night I finished reading a book, but while reading, the experience of jumping was constantly on my mind. . . .

# SUMMARY

A theoretical model for the measurement of approach-avoidance conflict was described, and predictions from it were evaluated by examining inexperienced parachutists at two time intervals from a jump. The model is a modification of Miller's model for approach-avoidance conflict. Gradients of drive were represented as a function of a stimulus and a time dimension. From the basic model of drive strength, models of activation and net approach increment were derived. It was assumed that the former could be measured by GSR and by performance deficit, and the latter by verbal content in response to a word association test with a built-in stimulus dimension. It was predicted that parachutists would produce gradients for GSR and for measures of performance deficit, and that the gradients would be higher and steeper on the day of a jump than at a time 2 weeks from a jump. It was further predicted that parachutists on the day of a jump would produce an increased number of goal relevant responses to the lower end of the stimulus dimension and a decreased number of such responses to the upper end of the dimension relative to their performance 2 weeks from a jump.

Eight novice sport-parachutists were tested first on the day of a jump and again 2 weeks later; eight others were tested first 2 weeks before a jump and again on the day of a jump. Sixteen nonparachutists served as controls. The tests consisted of parallel forms of a word association test with four levels of relevance to parachuting and with a group of words related to anxiety.

The major findings were as follows:

1. All 16 parachutists on both the day of a jump and the control day, and no control subjects, produced gradients of GSR as a function of the stimulus dimension. For all 16 parachutists, the gradients were higher and steeper on the day of a jump than on a control day.

2. Anxiety words elicited an increase in GSR from controls as well as parachutists. The increase was greatest for parachutists on the day of a jump, next for parachutists 2 weeks from a jump, and least for controls.

3. Parachutists produced smaller GSRs than nonparachutists to neutral words.

4. The results on reaction time, a measure of performance deficit, were similar to those for GSR.

5. Parachutists on the day of a jump exhibited a general perceptual deficit for neutral words, a greater deficit for anxiety words, and a relative sensitivity for parachute related words. The same results were obtained when parachutists were used as their own controls as when they were compared to nonparachutists. The findings can neither be accounted for by general frequency of or familiarity with words, or by inter- or intraserial set effects.

6. Parachutists on the day of a jump produced more parachute related responses than parachutists on a control day, who in turn produced more such responses than nonparachutists.

7. Three experienced jumpers on the day of a jump all produced a different form of curve for GSR and reaction time than inexperienced jumpers. The curve was similar up to a midpoint in the dimension, after which there was a sharp drop for the experienced jumpers and a continuous increase for the inexperienced jumpers.

It was concluded that the following three effects appear to be useful indicators of conflict: an increase in activation along a stimulus dimension, as indicated by a physiological measure, such as GSR; an increase in performance deficit along a stimulus dimension, as indicated by a formal (noncontent) measure, such as reaction time; and selective approach and avoidance, as indicated by perceptual sensitization and perceptual deficit, or by content of verbal response.



## REFERENCES

- EPSTEIN, S., & SMITH, R. Thematic apperception as a measure of the hunger drive. *J. proj. Tech.*, 1956, 20, 372-384.
- HATHAWAY, S. R. A comparative study of the psychogalvanic and association time measures: A new psychogalvanic apparatus. *J. appl. Psychol.*, 1929, 13, 632-646.
- HUNT, W. A., & LANDIS, C. Word association, reaction time, and the magnitude of the GSR. *Amer. J. Psychol.*, 1935, 47, 143-145.
- LANIER, L. H. An experimental study of "affective conflict." *J. Psychol.*, 1941, 11, 199-217.
- LURIA, R. *The nature of human conflicts*. New York: Liveright, 1932.
- MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, 66, 367-386.
- MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, 43, 155-178.
- MILLER, N. E. Comments on theoretical models illustrated by the development of a theory of conflict behavior. *J. Pers.*, 1951, 20, 82-100.
- MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science*. New York: McGraw-Hill, 1959. Pp. 196-292.
- MURRAY, E. J., & BIRKUN, M. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.*, 1955, 51, 47-56.
- PETERSON, F., & JUNG, C. G. Psychological investigations with the galvanometer and pneumograph in normal and insane individuals. *Brain*, 1907, 30, 211.
- WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

(Received January 16, 1961)

## MEDIATION AND ASSOCIATIVE FACILITATION IN NEUROTIC, PSYCHOTIC, AND NORMAL SUBJECTS<sup>1</sup>

PETER J. LANG AND KENNETH LUOTO

*University of Pittsburgh*

The mediation hypothesis assumes that the learning of a new response to a given stimulus may be facilitated if an association exists between this new response and some previously learned response to that stimulus. There is little general agreement as to the exact nature of the process (*rg.* MPS, cell assembly), but it seems impossible to account for complex behavior without a mediation construct. Osgood (1953) assumes mediation in his discussion of thinking. For Dollard and Miller (1950) verbal labels mediate emotional responses in psychotherapy. Goss (1958), Griffith, Spitz, and Lipman (1959) suggest that verbal mediators are fundamental to concept formation. For many, "the mediation postulate is made to carry the burden of insight, problem solving, and efficiency in transfer situations" (Bugelski, 1956).

Despite its importance to theories of thinking and verbal behavior, direct evidence for an implicit mediation process in verbal learning rests mainly with studies of college students (Bugelski & Scharlock, 1952; Kitao, 1960; Russell & Storms, 1955). The present research proposes to test the mediation hypothesis with an adult, noncollege sample of normal subjects, and to examine differences between these subjects and patients displaying behavior pathology. Two subgroups of the latter class, anxiety neurotics and schizophrenics, are of particular interest.

In a discussion of the differential performance of high and low anxious subjects in a paired-associate task, Spence (1958) suggests that the former will perform at a higher level if "the associative connections between each stimulus word and the nonpaired response words are lower than that to the paired word" (p. 139). These conditions are obtained in a

list, such as the one employed in the present research, which is composed of both pairs assumed to be related by a mediating term and pairs not so selected. Thus, we might anticipate that anxious subjects would show greater relative benefit from the mediational connection than "normal" subjects.

However, a contrary prediction is also applicable to the present research. In this experiment, the subject learns two paired-associate lists. Only the response terms are altered in the second list. Thus, in List II, the correct response is at the outset weaker than a potential intruder. Given approximately these conditions, Spence (1958) suggests that "the stronger the drive, the greater will be the percent choice of the wrong response, or, in other words, the poorer will be their performance at this initial stage" (p. 138). Although this prediction is relevant to the entire list, the conditions are especially enhanced for the nonmediation pairs. The list was constructed in such a way that, for these pairs, each stimulus member may be more closely related to another response term in the list than the one the subject is required to learn as part of the pair. Thus, considerable interference might be anticipated, retarding particularly the performance of anxious subjects on this part of the list. The present research will help determine the relevance of these drive theory predictions to mediation processes in verbal learning.

Since Bleuler (1950) disturbance of association has been considered a primary symptom of schizophrenia. This disturbance has been variously interpreted as a disorder of conceptual ability (Hanfmann & Kasanin, 1942), abstract behavior (Goldstein, 1939), or "looseness of association" (Meadow, Greenblatt, & Solomon, 1953). All of these imply either a defect of the mediational process in psychosis, or at least, a fundamental difference in the manner of its functioning in the

<sup>1</sup>This investigation was supported by United States Public Health Service Research Grant M-3567, Associative Transfer and Interference in Schizophrenia.



normal and schizophrenic patient. It would, thus, seem important to determine if the mediational effect in paired-associate learning can be demonstrated in the latter population, as it has been with college students.

Furthermore, Mednick (1958) has suggested that the thinking disorder of schizophrenics is a function of current anxiety, or represents the residual of a chronic anxiety state. His predictions of schizophrenic behavior in the setting under consideration parallel those of Spence for high drive subjects, with the qualification that the effects might be more extreme for the psychotic.

## METHOD

### *Experimental Design*

The design of this experiment was similar to that employed in their studies of college subjects by Russell and Storms (1955) and Bugelski and Scharlock (1952). In the latter experiment, subjects learned three lists of paired associations: List I established A-B associations; List II, B-C associations; List III tested the mediation effect by comparing the learning of A-C with A-X pairs.

Russell and Storms (1955) simplified the design by eliminating the learning of List II. In their study the B terms of List I were all items taken from the Kent-Rosanoff Word Association Test. The response terms used in the test list were associates of the B terms as defined by the normative data of the test.

The present design represents a further improvement, necessitated by the employment of pathological populations. Subjects were first administered a part of the Kent-Rosanoff Word Association Test, which contained among the 33 items administered the 10 B terms to be used in this experiment. The unique associates of each subject were subsequently used as the response terms in the second list.

### *Construction of the Paired-Associate Lists*

The first list learned by all subjects was the 10-pair nonsense syllable word list used as a first list by Russell and Storms (1955). The second list was composed of the same stimulus terms used in List I. Half of these stimuli were paired with response words that were mediated associations; the other five pairs (neutral) were made up of unrelated stimulus and response terms. However, the specific word used as responses for both mediated and neutral pairs varied systematically from subject to subject. For the first experimental subject, the first five response terms were the subject's unique associates to the List I response terms (taken earlier from the Word Association Test). The subject's own associations were also used for the second five response terms. However, they were as-

signed to stimuli randomly, with the restriction that each stimulus was assigned a response term unrelated to its response of List I. List II for subsequent subjects was constructed in a similar manner, except that the five stimuli receiving mediated and the five receiving unrelated (neutral) response terms were alternated through the sample. Thus, differential learning of mediated and nonmediated pairs could not be ascribed to the ease of learning a particular set of nonsense syllables or the associates of any particular set of response terms.

### *Subjects*

The design described above was replicated with three different subject populations. The first group was composed of 15 patients convalescing on the surgical wards of the Oakland Veterans Administration Hospital, Pittsburgh, Pennsylvania. They presented no history of psychiatric disorder; patients with physical complaints frequently termed "psychosomatic"—stomach ulcers, low back pains, asthma—were also excluded from this group. Their mean age was 32.13 years. The average Wechsler Vocabulary score was 23.53, and the *MA* scale stanine was 4.73.

The second group consisted of 15 patients from neuropsychiatric wards of the Oakland Veterans Administration Hospital. All these subjects carried a diagnosis of "anxiety reaction" or "anxiety neurosis" and were judged to be anxious by attendant personnel. The mean *MA* scale stanine for this group was 8.2. They displayed no psychotic symptoms nor revealed any history of psychosis. The mean age of the anxious subjects was 35.13 years. Their vocabulary score was 26.06.

Fifteen chronic schizophrenic patients in partial remission made up the third group. They were seen at the Leech Farm Veterans Administration Hospital, Pittsburgh, Pennsylvania. These subjects had no history of neurological involvement or primary mental deficiency. Their vocabulary score was 29.73. The average age of this group was 30.10 years.

In general, all male patients between the ages of 20 and 41, who met the diagnostic criteria of the appropriate subgroup, had an equal opportunity to participate in the experiment. However, this condition was overridden by either of the following factors: refusal to participate; the subject judged to be "out of contact and untestable" after consultation between the experimenter and the ward attendant. Furthermore, if a patient had learned less than five pairs after 30 presentations of the entire list, it was held that the task was too difficult to be meaningful for this subject, and he was dropped from the group. Six subjects were eliminated for this reason.

### *Procedure*

The subject was first administered the Word Association Test, followed by the Wechsler-Bellevue Vocabulary list. Subsequently the subject learned two

TABLE 1

MEAN NUMBER OF TRIALS REQUIRED TO ACHIEVE CRITERION FOR MEDIATED PAIRS, NEUTRAL PAIRS, AND THE TOTAL LIST

Diagnosis	List	Mediation	Neutral	Total
Normal	I	14.31	14.53	14.42
	II	9.67	11.53	10.58
Anxiety	I	12.76	13.75	13.25
	II	9.25	12.35	10.75
Schizophrenia	I	14.39	13.95	14.17
	II	11.64	14.93	13.30

lists of nonsense syllable-word pairs, with a 15-minute interval between lists. The pairs were typed on individual 3" x 5" cards and viewed by the subject through an opening in a black, plywood panel. Behind the panel a motor driven, slotted screen was arranged to expose the stimulus half of each pair for 2 seconds and the response half for an additional 3 seconds, with 3.5 seconds between pairs (Bugelski & Scharlock, 1952).

The subject learned to anticipate each correct response to a criterion of four successive, perfect trials. As the criterion for each pair was met, that pair was dropped from the list to prevent over-learning. The cards were shuffled between trials to eliminate serial learning. The entire task was usually completed in less than an hour and a half, including the interval between lists during which the subject relaxed with a magazine. Following the experimental session, the neurotic and nonpsychiatric patients were administered a short form of the Taylor MA scale (Bendig, 1956).

All correct anticipations and errors were tabulated for each subject, for both Lists I and II. The measures used in the final analysis were: the mean number of trials required to reach criterion on all pairs in List I, and separately for List II; the mean number of trials required to reach criterion on all five mediated pairs, and separately on all five neutral pairs, for both Lists I and II. For List I "mediated pairs" refers to those stimuli which acquired the mediated response term in List II.

These measures were used because they seemed to provide the least biased estimate of overall performance. However, they offer no information about the rate or progress of learning. Therefore, the number of trials required to reach criterion on each of the 10 successive pairs in List I and II were also tabulated.

In discussing the relevance of drive theory to paired-associate learning, Spence (1958) makes the following point: "... the theory of paired-associates learning has as yet not been developed sufficiently to predict what will happen beyond the first few trials, and it would have been more appropriate, as far as implications for our drive theory are concerned, if we had used at most only the data from

the first four or five trials" (p. 139). In accordance with this admonishment, the number of correct responses made in the first five trials of List II were separately recorded for the total list, the mediated pairs, and the neutral pairs.

## RESULTS

Despite differences in diagnosis, no difference in initial learning between subgroups was observed. A variance analysis of the mean number of trials required to learn List I resulted in an  $F$  less than 1.0. Criterion scores achieved by each patient group on both Lists I and II are presented in Table 1.

The statistic used to evaluate all these data was the analysis of variance for repeated measures, in which there are two variables within-subjects (list and mediation) and one between-subjects (hospital diagnosis). Because specific hypotheses were raised concerning differences between psychotic populations and between anxious (high drive) and non-anxious (low drive) subjects, these com-

TABLE 2

ANALYSIS OF VARIANCE OF MEAN TRIALS TO CRITERION FOR DIAGNOSIS, LIST, AND MEDIATION

Source	df	MS	F
Diagnosis	2	46.04	
Psychosis	1	85.08	1.10
(Schizophrenic vs. Anxiety and Normal)			
Anxiety	1	7.00	—
(Anxiety vs. Normal)			
Pooled subjects with ingroups	42	77.33	
Lists	1	255.85	11.80**
Diagnosis X Lists	2	32.47	1.49
Psychosis X Lists	1	50.93	2.34
Anxiety X Lists	1	14.01	—
Pooled subjects X Lists within diagnosis	42	21.68	
Mediation	1	101.85	11.86**
Diagnosis X Mediation	2	3.77	—
Psychosis X Mediation	1	.13	—
Anxiety X Mediation	1	7.41	—
Pooled subjects X Mediation within diagnosis	42	8.59	
Lists X Mediation	1	69.94	8.16*
Diagnosis X Lists X Mediation	2	4.53	—
Psychosis X Lists X Mediation	1	8.64	1.008
Anxiety X Lists X Mediation	1	.41	—
Pooled subjects X Lists X Mediation within diagnosis	42	8.57	

Note.—The between diagnosis variance has been further subdivided into two components called Psychosis, the schizophrenic group vs. the anxiety and normal groups; and Anxiety, the anxiety group vs. the normal group.

\*  $p < .02$ .  
\*\*  $p < .01$ .



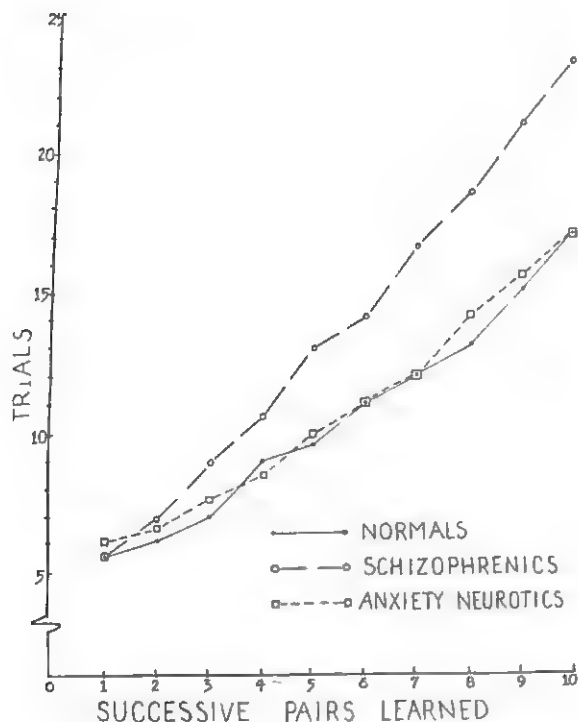
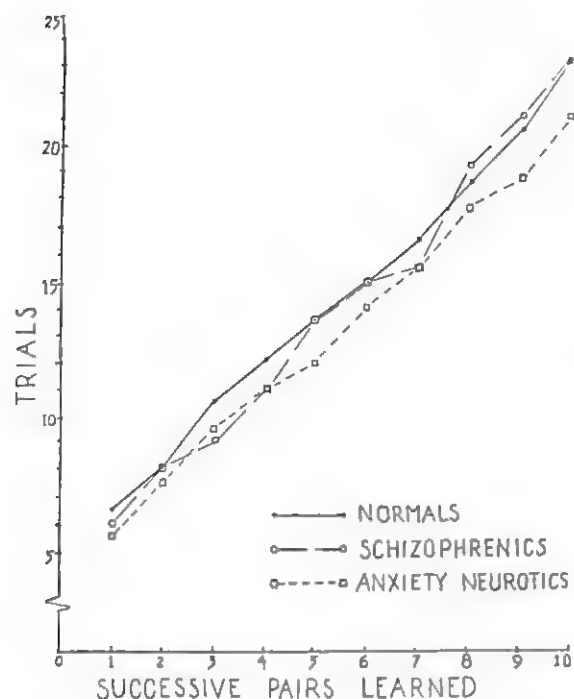


FIG. 1. Mean number of trials required to learn successive pairs for List I and List II.

ponents of the between-groups variance were separately analyzed.<sup>2</sup>

The results of this analysis are presented in

<sup>2</sup> The authors wish to thank A. W. Bendig who designed this analysis and Maryl Conwisher who assisted with the computation.

Table 2. The Lists  $\times$  Mediation  $F$  value is significant beyond the .02 level, providing clearcut evidence that the mediation process extends across diagnostic groups. Furthermore, there is no indication in this analysis of a differential mediation effect for either anxious or psychotic subjects. Neither the overall Diagnosis  $\times$  Lists  $\times$  Mediation interaction or the separate component  $F$ s approached accepted significance levels.

In general all groups showed some associative facilitation, improving their scores from List I to List II. The analysis of overall differences between lists yielded an  $F$  of 11.80 ( $p < .01$ ). Although the main Group  $\times$  Lists interaction did not approach significance, the psychotic vs. nonpsychotic component attained a  $p$  value close to .10. The direction of this trend is revealed in Table 1 where it will be noted that the degree of improvement from List I to List II decreases progressively from a difference in means of 3.84 for the psychiatrically normal to less than a trial (.87) for the schizophrenic patients.

Differences between psychotic and nonpsychotic patients in associative facilitation are clearly revealed when the course of learning is examined. Figure 1 shows the progress of learning for the three diagnostic groups on both List I and List II. Inspection reveals no differences between groups for List I. This was confirmed by an analysis of variance in which the Diagnosis  $\times$  Successive Pairs interaction  $F$  was less than 1.0. However, for List II the schizophrenic subjects displayed a tendency to take an increasing number of trials than either anxious or normal patients to learn successive pairs. The graphic picture was confirmed by an analysis of variance: the Diagnosis  $\times$  Pairs  $F = 2.33$ ,  $p < .01$ .

The analysis of the number of correct responses in the first five trials of List II was not undertaken directly, because of the potentially confounding effect of List I learning. It was decided to remove this variable through covariance analysis. However, the measure of List I learning to be used presented a problem: the correlation between the first five trials of List I and List II was low and not significant. Fortunately, mean trials to reach criterion on List I and the relevant

TABLE 3  
MEAN NUMBER OF CORRECT RESPONSES IN THE FIRST FIVE TRIALS OF LIST II

Diagnosis	Mediation		Neutral		Total	
	Actual <sup>a</sup>	Adjusted <sup>b</sup>	Actual <sup>a</sup>	Adjusted <sup>b</sup>	Actual <sup>a</sup>	Adjusted <sup>b</sup>
Normal	11.27	11.57	6.87	7.00	18.13	18.56
Anxiety	9.27	8.61	5.80	5.70	15.07	14.43
Schizophrenia	8.87	9.17	4.27	4.23	13.13	13.13

<sup>a</sup> Refers to the actual mean.

<sup>b</sup> Indicates List II mean scores adjusted for the regression of the respective List I mediation, neutral, and total criterion scores.

Y variable yielded an  $r$  of  $-.60$ . Requirements for homogeneity of regression were met with this measure, and it became the X variable in the subsequent analyses.

Three separate analyses of covariance were performed on the following scores: total correct responses, correct mediated responses, and correct neutral responses. The actual and adjusted means for each variable are presented in Table 3. The results of the analyses, presented in Table 4, reveal differences between patient populations in total correct responses. A test of normal and schizophrenic means yielded a  $t$  of 2.48,  $p$  of .02. A similar test of differences between anxious and non-psychiatric patients resulted in a  $t$  of 1.95,  $p$  of .06.

The  $F$  of 3.17 for correct neutral terms also warrants a test of individual means. In this case the differences between normal and schizophrenic subjects yielded a  $t$  of 2.14,  $p < .05$ . The difference between the non-psychiatric and anxiety group means was not significant.

TABLE 4  
ANALYSIS OF COVARIANCE OF THE FIRST FIVE TRIALS OF LIST II

Source	df	MS	F
Mediated pairs			
Diagnosis	2	32.73	2.11
Error	41	15.53	
Neutral pairs			
Diagnosis	2	28.81	3.17*
Error	41	9.08	
Total list			
Diagnosis	2	113.03	3.39**
Error	41	33.30	

\*  $p < .06$ .

\*\*  $p < .05$ .

## DISCUSSION

The results of this experiment clearly demonstrate the facilitating effect of verbal mediation on learning in nonstudent populations. They also suggest that the mediation process is not fundamentally impaired in schizophrenia, and that theories resting on this assumption may be in error.

The findings relevant to the hypotheses of Mednick (1958) and Spence (1958) are complex, and require careful review. First of all, as Spence predicts, differences between groups appeared in the analysis of the "first few trials" of List II, rather than in the evaluation of mean trials to criterion.<sup>3</sup> It would appear that, despite the fact of overall associative facilitation, old response terms interfered differentially in the acquisition by the separate subgroups of the new pairs. The effect of these intruders was most marked for the schizophrenic subjects, as Mednick proposes, although the anxious subjects also showed impaired performance relative to normals. The prediction of an interference ef-

<sup>3</sup> Nevertheless, inspection of the individual criterion scores for the three groups was suggestive. Only one member of the normal group failed to achieve criterion on the mediated pairs of List II in fewer trials than on List I. Both the anxious subjects and the schizophrenic patients behaved comparably on the mediated pairs: only two subjects from each of these groups gave similar evidence of negative transfer. However, the neutral pairs yielded quite different results: three normal, six anxiety neurotics, and nine schizophrenic patients (subgroups  $n = 15$ ) learned List II in more trials than were needed to learn List I. This finding parallels the subgroup differences noted on the first few trials of List II. It also implies, at least in the case of neutral pairs, that List I learning interfered with List II learning, and furthermore that the interference effect was positively related to an increase in behavior pathology.



fect for high drive subjects on neutral pairs was realized for the schizophrenic group but not for the anxious subjects.

The fact that no difference between groups was apparent for mediated pairs is comprehensible in Spence's (1958) terms, if we consider the fact that two opposing factors are held to be operating. Although in this case the associative connections between the stimulus and its paired word were initially greater than with the other response terms (suggesting an advantage for anxious and schizophrenic subjects); subjects were also being asked to learn a response initially weaker than a potential intruder (i.e., the response learned 15 minutes earlier). It is reasonable to assume that advantage and disadvantage for high drive subjects cancel each other out, resulting in no differences between groups.<sup>4</sup>

The difference in learning curves between psychotic and nonpsychotic subjects is not readily integrated with the other findings. Spence's (1958) drive hypothesis suggests that differences should appear early rather than late in learning. The picture of increasing deviation with successive pairs implies the effects of a second variable.

The possibility that curve differences are an artifact of the experimental design was explored. The response terms for List II varied from subject to subject, and there might have been systematic differences between groups. As would be anticipated (Moran, 1953) the schizophrenic subjects tended to report less common associations than did the normal or anxious subjects. It is conceivable that these were more difficult to learn, and this affected the rate at which the subject reached criterion on successive pairs. In point of fact, this variable might have contributed to subgroup differences on the other measures already discussed. However, a Pearson  $r$  (sample  $N = 18$ ) between the number of pairs correct in List II and an

index of deviation based on the Thorndike-Lorge word count was insignificant and in the direction which contradicts this hypothesis.

Two possibilities remain which could account for the findings. The schizophrenics might have suffered from unique fatigue effects or flagging motivation. However, this hypothesis seems difficult to accept in the light of their failure to show this phenomenon on List I, and the fact that List II learning was slightly more rapid, and followed a 15-minute rest period. A second, more likely possibility was suggested by the behavior of the psychotic group during learning. Late in learning, these subjects frequently responded to a stimulus syllable with a response term which had already achieved criterion and been dropped from the list. This happened rarely with the other groups. For these latter subjects, when a response was eliminated its status as a potential intruder was markedly reduced. The schizophrenics, on the other hand, behaved as if nearly all the competitors remained active, if not phenomenally present, until the very end of the list. In List I the effects might be minimal. However, in List II, where a neutral stimulus was always more closely related to some response other than the one with which it was paired, and intruders from List I could reassert themselves, this might have greatly retarded the learning of the final few pairs.

The above interpretation raises the question of the extent to which subjects were "conscious" of mediators or interfering terms during the learning. Previous investigators (Bugelski & Scharlock, 1952; Foley & Cofer, 1943) have expressed considerable interest in this issue; however, an adequate method of determining awareness has not been developed.

In the present research, the last 5 subjects seen in each diagnostic group were closely questioned at the end of the experimental session, concerning their thoughts and impressions during learning. Only 2 of the 15 subjects interviewed suggested that they knew the List II response terms to be their own associations. Eight of the 15 said that they were aware of a relationship between the two lists, although the connection was usually

<sup>4</sup> It is also possible that differences failed to appear because the psychiatrically disturbed patients offered unique associations which were systematically different in their ability to mediate generalization. It might be hypothesized that these associations were less stable, less able to mediate generalization, and the effect of drive variables was simply to raise the learning of anxious subjects to the same level of proficiency as that displayed by the normal patients.

poorly understood. All 7 who were totally unaware of facilitating or interfering associations displayed the positive effects of the mediated connections in List II learning. This latter finding is consistent with previous research, which suggests that awareness is not necessary to verbal mediation.

The distribution of awareness by diagnostic groups is of some interest. Four schizophrenics, two anxious, and one normal subject were unaware of connecting associations. Thus, insight seems to follow the pattern of interference effects. It is possible that while awareness does little to facilitate mediated association, it may be an important factor in keeping out intruders. Subjects who are aware that a compelling connection is nevertheless wrong, are conceivably in a better position to ignore that association than a subject who fails to perceive any relationship at all.

In discussing the performance of schizophrenic subjects on a psychomotor task, the senior author suggested that deficit in performance was attributable to a failure of "set" or "self-instruction"—an inability to discriminate relevant cues during performance and reprogram the organism to take account of them (Cavanaugh, Cohen, & Lang, 1960; Lang, 1959). The lack of awareness displayed by the schizophrenics on the mediation task suggests a similar process, which may have contributed both to retardation in the course of learning and the deficit noted on neutral and total pairs in the first five trials of List II.

In the absence of a sampling design attesting to the representativeness of our samples of schizophrenic and other populations, our results bear primarily on the mediation process in association rather than on differences between types of subjects. However, the present findings strongly suggest that special transfer and interference phenomena are important contributors to the symptom picture in chronic schizophrenia.

# SUMMARY

Schizophrenics, anxiety neurotics, and psychiatrically normal patients ( $N = 45$ ) learned two lists of 10, syllable-word paired associates. Stimulus terms were identical on Lists

I and II. Half of the List II pairs had response terms which were associates of their List I responses; the response terms of the other five (neutral) pairs were not related to their stimuli by any such mediating association.

The entire population showed improvement from List I to List II, and greater improvement for mediated than for neutral pairs. Schizophrenic subjects took progressively more trials than the other two groups to learn successive pairs in List II. Both the anxiety neurotics and the schizophrenics had fewer correct responses from the total list in the first few trials of List II than did normal subjects. Schizophrenics also responded with fewer correct neutral pairs than the normals.

# REFERENCES

- BENDIG, A. W. The development of a short form of the manifest anxiety scale. *J. consult. Psychol.*, 1956, 20, 384.
- BLEULER, E. *Dementia praecox or the group of schizophrenics*. New York: International Univer. Press, 1950.
- BUGELSKI, B. R. *The psychology of learning*. New York: Holt, 1956.
- BUGELSKI, B. R., & SCHARLOCK, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, 44, 334-338.
- CAVANAUGH, D. K., COHEN, W., & LANG, P. J. The effect of "social censure" and "social approval" on the psychomotor performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1960, 60, 213-218.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy: An analysis in terms of learning, thinking and culture*. New York: McGraw-Hill, 1950.
- FOLEY, J. P., & COFER, C. N. Mediated generalization and the interpretation of behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, 32, 168-175.
- GOLDSTEIN, K. The significance of special mental tests for diagnosis and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, 96, 575-588.
- GROSS, A. E. Mediating responses and concept formation. *Amer. Psychologist*, 1958, 13, 391. (Abstract)
- GRIFFITH, B. C., SPITZ, H. H., & LIPMAN, R. S. Verbal mediation and concept formation in retarded and normal subjects. *J. exp. Psychol.*, 1959, 58, 247-251.
- HANFMANN, EUGENIA, & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. ment. dis. Monogr.*, 1942, No. 67.
- KITAO, N. Effects of mediated association on paired-associate learning: I. Facilitation effect and interference effect. *Jap. J. Psychol.*, 1960, 31, 1-6.



- LANG, P. J. The effect of aversive stimuli on reaction time in schizophrenia. *J. abnorm. soc. Psychol.*, 1959, 59, 263-268.
- MEADOW, A., GREENBLATT, M., & SOLOMON, H. C. "Looseness of association" and impairment in abstraction in schizophrenia. *J. nerv. ment. Dis.*, 1953, 118, 27-35.
- MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychol. Bull.*, 1958, 55, 316-327.
- MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, 67(20, Whole No. 370).
- OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.
- RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, 49, 287-293.
- SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, 13, 131-141.

(Received January 23, 1961)

## A DISSONANCE THEORY APPROACH TO DEFENSIVE PROJECTION<sup>1</sup>

DANA BRAMEL

*University of Minnesota*

Ego defensive processes, as discussed in psychoanalytic theory, often seem to bear some resemblance to the cognitive changes dealt with in Festinger's (1957) dissonance theory. This observation has led to a comparison of the two theories and to consideration of the possibility that certain of the Freudian defense mechanisms might occur in response to dissonance. Especially interesting from the point of view of social psychology is the concept of projection, since it clearly has implications for interpersonal relations.

Consider those situations, described in psychoanalytic theory, in which the individual's perception of some aspect of himself is contrary to his internalized standards of right and wrong (the superego). According to the theory the perception of this information arouses fear of punishment, perhaps especially a fear of painful guilt feelings (Fenichel, 1945). In order to avoid further anxiety and guilt feelings, the ego is said to initiate defensive measures.

In order to determine the relevance of dissonance theory to this phenomenon, one must ask whether dissonant relations would be expected to exist among the cognitions involved. Imagine, for example, a person who considers homosexuality a bad and disgusting thing; on some occasion he is suddenly exposed to information strongly implying that he has homosexual tendencies. According to classical psychoanalytic theory, the crucial relation is the conflict between the information or impulse and the demands of the superego. Is the new cognition—that one has

homosexual tendencies—necessarily dissonant with one's belief that such tendencies are bad and that one should not have them? The answer is no. People who conceive of themselves as possessing a mixture of good and bad traits, or a preponderance of bad traits, would not generally expect that new information would be favorable to themselves or consistent with superego standards. It follows that a discrepancy between the new information and superego standards, although threatening in the psychoanalytic sense, would not necessarily be dissonant.

This is not to say, however, that dissonance would be completely absent from the cognition of the person who recognizes that he is not perfect. For example, when he discovers he has homosexual tendencies, this knowledge may be dissonant with his specific belief that he is really quite masculine, even though it is not dissonant with his conviction that homosexuality is a bad thing.

The point can be clarified by a hypothetical example. Imagine two people, A and B. Both consider homosexuality a very bad thing, and both believe they are quite lacking in such motivation. A believes he is an extremely fine person in general; B sees himself as possessing almost no favorable characteristics. Both are then confronted with information that they have strong homosexual tendencies. For both this information is contrary to superego standards and dissonant with their belief that they are not homosexual. However, for A the information is also strongly dissonant with his belief that he is a nearly perfect person. For B, on the other hand, the information is quite consonant with his belief that he is a failure. The new information produces more dissonance for the person with high self-esteem, even though the conflict with the superego is substantially the same for the two people. This emphasis upon the actual self-concept in the dissonance theory ap-

<sup>1</sup> This article is based upon a doctoral dissertation submitted to Stanford University. The author wishes to express his indebtedness to Leon Festinger for valuable assistance throughout the research. Advice and criticism from Robert R. Sears, Nathan Maccoby, Quinn McNemar, and Stanley Schachter are gratefully acknowledged. Special thanks are also due Alvin I. Haimson for assistance in conducting the experiment.

proach reveals a difference in focus of the two theories.

Can dissonance involving the self-concept be reduced by projecting the offending trait onto other people? Perhaps the most effective mode of dissonance reduction would be to deny the implications of the information. Let us assume, however, that the information is so unambiguous that successful denial is not possible for the person. He is compelled to ascribe the undesirable trait to himself. Under these circumstances, attributing the trait to other persons might reduce dissonance in several ways. By attributing it to respected people, the projector may enable himself to re-evaluate the trait. If respected persons possess it, then perhaps it is not so bad a thing after all. Then possession of the trait would not be contradictory to a favorable level of self-esteem. Another possibility is that the person may attribute the trait to members of his reference or comparison group (Festinger, 1954). In this way he could convince himself that he does not deviate from the persons with whom he ordinarily compares himself. If he is only average in his possession of the trait, then subjectively his favorable level of self-esteem is not so strongly negated.

These possibilities suggest that indeed projection may be used as a means of reducing dissonance. There are several studies in the literature (for example, Backman & Secord, 1959 unpublished; Murstein, 1956; Wright, 1942) which are specifically relevant and show positive results, but all leave certain important issues unresolved. Consequently, the experiment reported here was conducted to test whether projection occurs in response to dissonance and to throw some light on the particular ways in which this attribution may reduce dissonance.

The hypotheses to be tested were these:

1. If a person is exposed to information strongly implying that he possesses an undesirable characteristic, he is more likely to attribute the trait to others if the information is dissonant with his level of self-esteem; the greater the dissonance, the more likely it is that projection will occur.

2. If a person is compelled to ascribe an undesirable and dissonant characteristic to

himself, he will be motivated to attribute the characteristic to favorably evaluated persons and/or to persons with whom he ordinarily compares himself.

## METHOD

*Overview.* Subjects in the Favorable condition received falsified personality test results aimed at temporarily increasing their general level of self-esteem; subjects in the Unfavorable condition received parallel information intended to lower their general self-esteem. Subsequently, all subjects were privately exposed to further falsified information of an undesirable nature about themselves. It was hypothesized that this information, being more dissonant with the self-concept of subjects in the Favorable condition, would lead to more projection in that condition. Attribution was measured by asking each subject to rate another subject with whom he was paired.

*First session.* Each subject who signed up for the experiment appeared individually for the first session. He was told that the first part of the experiment was designed to discover what kinds of people had insight into themselves. He was asked to take a number of personality tests, which, he was told, would be carefully and confidentially analyzed by three members of the clinical psychology staff. He was informed that, after the tests were scored, he would learn the results in an interview, during which time his self-insight was to be measured. Among the tests included were the Taylor Manifest Anxiety scale, the *F*, *K*, and *L* scales from the MMPI, and an adjective checklist self-concept measure.

At the end of the hour, the subject was told that the second session would also include a measure of his ability to judge the personality of another person on the basis of a first impression.

*Second session.* On the basis of the self-concept measure subjects were paired for the second session by matching their level of self-esteem and their concept of their own masculinity.

At the beginning of the second session the two unacquainted subjects scheduled for the hour were introduced to each other. In order to aid subjects in forming an impression of each other, the experimenter asked each in turn (in the presence of the other) a set of questions about himself and his attitudes toward certain current events. At the conclusion of this meeting the subjects were separated and interviewed privately regarding the results of their personality tests.

Unknown to the subject, the "results" which he received had been prepared with no reference whatsoever to his actual test performance. There were only two test reports used in the experiment, one very favorable, and the other very unfavorable. The reports covered the personality "dimensions" of creativity, hostility, egocentricity, and over-all maturity. Each section of the report gave a rather detailed discussion of the test results bearing upon



the particular dimension. The tone was objective and the general favorability was very consistent throughout the report. The two reports were very similar in form, but the specific contents were directly opposite in implication.

After having been assigned randomly to his experimental condition, the subject was read the report in private by the experimenter, and its discrepancies from the subject's present self-concept were explicitly pointed out by the experimenter. The report was finished in approximately 20 minutes. In each pair of subjects, one was assigned to the Favorable condition, one to the Unfavorable condition. Two interviewers were used for this part of the experiment, alternating between the two conditions.

Following the test report, the two subjects were brought together in another room, where they expected to make some personality judgments about each other. Each was then given a questionnaire consisting of 11 polar adjective seven-point scales, to be used to rate the other person. An over-all favorability score could be computed across the scales. Examples were masculine-feminine, friendly-hostile, competent-incompetent, and mature-immature. A self-concept measure followed, consisting of 16 polar adjective scales similar to those included in the prior rating of the other person. These were selected partially to tap dimensions covered in the personality reports, and to serve as a check on the manipulation of self-esteem. As in the previous set, they included the item, masculine-feminine, and could be summated to provide a general favorability score. Emphasis was placed upon the anonymity of the questionnaires and upon the earnest request that the subject respond "as you see yourself, from your own point of view at the present time."

*Introducing the undesirable cognition.* It was expected that informing a male undergraduate that he has homosexual tendencies would be sufficiently dissonant under certain conditions to provoke defensive behavior. Care was taken to ensure that the degree of threat was not extreme and that no damaging effects would remain at the end of the experiment. These precautions will be discussed in more detail below.

At this point in the experiment, while making ratings of partner and self, the two subjects were seated along one side of a long table, separated by about 4 feet, both facing a projection screen 6 feet in front of them. On the table in front of each subject was a small plywood box containing a galvanometer dial facing him. Issuing from the box were two wires with electrodes on the ends. Each box, with its dial, was completely shielded from the other persons in the room. Thus, each subject perceived his apparatus immediately in front of him and could not see the other subject's apparatus.

Next, the experimenter read a set of instructions to set the stage for the undesirable cognitions about homosexuality. These instructions were largely of a deceptive nature. It was explained that this part of the experiment would be concerned with the

perception of sexual arousal. An elaborate explanation of the physiology of sexual arousal and the sensitive techniques for its measurement followed. Care was taken to distinguish the galvanometer response to sexual arousal from that commonly associated with anxiety reactions, by pointing out the unmistakable signs of the former. Considerable emphasis was placed on the unconscious nature of sexual arousal and the impossibility of exerting conscious control over its expression in the "psychogalvanic skin response." It was further explained that the experimenter was investigating the perception of homosexual rather than heterosexual arousal. The task set for the subject was to observe his own sexual arousal response on his galvanometer for each of a series of photographs of men which would be projected onto the screen. He was to record this figure on a page of a small anonymous booklet. After he had recorded his own arousal level for the particular picture on the screen, he was to make an estimate of the needle indication of the other subject's apparatus for the same photograph. All subjects were explicitly told that movements of the dial would indicate homosexual arousal to the photographs. As a precaution against excessive homosexual tendencies would consistently "go off the scale." Further, the anonymity and privacy of the situation were carefully spelled out, with the intention of convincing the subject that no one but he would know what his own responses had been.

Unknown to the subject, the supposed "psychogalvanic skin response apparatus" was not actually responding to changes in his own level of sexual arousal to the pictures. Rather, the galvanometers in each of the two boxes were controlled remotely by the experimenter. Concealed wires led from the galvanometers, in a direct current series circuit, to a calibrated variable resistor. Thus, the experimenter exerted complete control over the movements of the needles, which were identical for the two subjects. Each photograph had been assigned an "appropriate" needle reading in advance, so that those depicting handsome men in states of undress received more current than did those depicting unattractive and fully clothed men. Both subjects were, thus, led to believe that they were sexually aroused by certain pictures and not by others, according to a consistent pattern. Both subjects were confronted with exactly the same stimulus input at this point of the experiment.<sup>2</sup>

It was expected that the instructions would be so impressive to the subject that denial of the fact that homosexual arousal was being indicated would be very difficult. By closing off certain alternative avenues of dissonance reduction, such as the cognition that the apparatus was untrustworthy, it was intended that the situation would be conducive to the appearance of defensive projection. According to the hypothesis, subjects in the Favorable con-

<sup>2</sup> A similar experimental technique was independently devised by Harold Gerard (cf. Gerard, 1959; Gerard & Rabbie, 1960).

dition should experience considerable dissonance when observing their needle jump in response to photographs of attractive males. For subjects in the Unfavorable condition there would be many cognitions consonant with the new information concerning homosexuality, and not so many dissonant cognitions. For most subjects there would, no doubt, be some dissonance due to their prior belief that they are not homosexual, but the two conditions would not differ in this respect. Since the test reports contained no material concerning sexuality, subjects in the two conditions were not expected to differ in their concepts of their own masculinity or in their superego standards.

Fifteen photographs of men were used. Many of the men were almost entirely nude and had physiques somewhat more delicate and posed than those typically found in physical culture magazines. These photographs were chosen on the assumption that subjects might perceive them as being the type toward which homosexuals would be attracted.

*Measuring attribution.* It seemed that the most meaningful measure for testing the hypotheses would be a score representing the difference between the subject's own recorded score and his estimate of his partner's galvanic skin response. This should most accurately reflect the subject's comparison between himself and his partner. Therefore, a total score was computed for each subject, taking the algebraic sum of the differences between own and attributed scores across the 15 photographs. This summary score (*P* score) would be positive if the subject attributed (on the average) higher needle indications to his partner than to himself (i.e., attributed greater homosexual arousal). It would be zero if on the average there was no difference between own and attributed scores. It would be negative if the subject attributed lower needle indications to the other subject than to himself.

Following the threatening material, the subject responded to anonymous questions about his own and his partner's degree of possession of homosexual

tendencies, and about his attitude toward the "psychogalvanic skin response" as a measure of such tendencies.

A considerable amount of time at the end of the experiment was allocated to explaining the true nature of the study and demonstrating in detail that the personality reports and the apparatus were incapable of giving a correct evaluation of a person. The expression of relief which often followed the unveiling of the deceptions indicated that the manipulations had been effective. The necessity for the deceptions used in the experimental analysis of such delicate processes was carefully explained, and all questions were answered. Not until the subjects seemed quite restored and satisfied was the session ended. All available evidence indicates that the subjects considered the experiment interesting and worthy of their participation.

*Subjects.* All subjects in the experiment were undergraduate men registered in the introductory psychology course. Not all of those who took part in the first session were selected to finish the experiment. Those who scored very high on the Taylor Manifest Anxiety scale and at the same time very low on the defensiveness scale of the MMPI were excluded from the second session, since there was the possibility that the manipulations might be too threatening for them. Of the 98 subjects who participated in both sessions of the experiment, 14 were excluded from the analysis—8 for suspicion regarding the procedure, 3 for excessive age (over 30), and 3 for failure to obey the instructions. Of those excluded, 7 were in the Favorable and 7 in the Unfavorable condition.

## RESULTS

*Adequacy of experimental operations.* The major independent variable was the level of self-esteem, or the number of favorable self-referent cognitions. A checklist measure of self-esteem administered before the manipulations revealed no initial difference between the groups. The effectiveness of the personality reports was determined by comparing the Favorable and Unfavorable groups on self-esteem as measured by adjective scales shortly after the manipulation. Mean favorability scores could range from a low of 1 to a high of 7. The results are shown in Table 1. The mean for the Favorable group was significantly higher than for the Unfavorable group ( $t = 8.35, p < .001$ ).<sup>3</sup> We may infer, therefore, that the desired difference in self-esteem was successfully created by the fraudulent test reports.

Another important problem of experimental

TABLE 1

MEANS AND STANDARD DEVIATIONS OF VARIABLES  
MEASURED PRIOR TO INTRODUCTION OF  
THE UNDESIRABLE COGNITION

Variable	Favorable ( <i>n</i> = 42)	Unfavorable ( <i>n</i> = 42)
Before self-esteem manipulation		
Initial self-esteem (checklist)		
<i>M</i>	14.5	15.1
<i>SD</i>	3.6	3.3
After self-esteem manipulation		
Self-esteem (seven-point scales)		
<i>M</i>	5.52	4.20
<i>SD</i>	.54	.86
Favorability of rating of partner		
<i>M</i>	4.79	4.90
<i>SD</i>	.69	.68
Rating of masculinity of partner		
<i>M</i>	5.39	5.30
<i>SD</i>	1.11	1.32
Rating of own masculinity		
<i>M</i>	5.87	5.56
<i>SD</i>	1.00	1.02

<sup>3</sup> All reported significance levels are based upon two-tailed tests.



control had to do with the favorability of the subjects' ratings of each other prior to introduction of the cognitions concerning homosexuality. A score was calculated for each subject, taking the mean of his ratings of his partner (scored for favorability) across 10 polar adjective scales, excluding the item, masculine-feminine (considered separately below). The first half vs. second half reliability of the score was .57. Possible scores could range from 1 (very unfavorable) to 7 (very favorable). As shown in Table 1, there was no significant difference between the two conditions in favorability of rating of partner.

On the masculine-feminine scale, 1 indicated "very feminine" and 7 indicated "very masculine." Mean scores of the two groups did not differ on this scale, as shown in Table 1. Further, the groups did not differ significantly in their rating of their own "masculinity-femininity."

On the basis of these comparisons, it seems justifiable to conclude that the Favorable and Unfavorable groups did not differ regarding these possible artifactual effects of the self-esteem manipulations.

*Self-esteem and projection.* Before comparing the two experimental conditions on the attribution of homosexual arousal, let us check the reliability of the measure, the *P* score. For the first 23 subjects used in the experiment, the series of 15 photographs was repeated, yielding a set of 30 judgment situations for each subject. The discrepancies between his own recorded dial readings and his estimates of his partner's dial were summed separately for the first 15 and the second 15 exposures of the photographs. The correlation between these two sums (*P* scores) was .95. For subsequent analyses, only the *P* score for the first 15 photographs was used.

Subjects in the two conditions did not differ, on the average, in their own recorded scores. That is, they were equal in the accuracy with which they recorded their own needle indications. Therefore, the *P* scores, which were partially derived from the subjects' own recorded dial readings, could not differ between the two groups simply as a function of differences in own recorded scores.

For evidence concerning the relation between dissonance and projection, let us look

TABLE 2  
MEANS AND STANDARD DEVIATIONS OF ATTRIBUTION  
MEASURED AFTER MANIPULATION  
OF SELF-ESTEEM

Attribution	Favorable ( <i>n</i> = 42)	Unfavorable ( <i>n</i> = 42)
Raw <i>P</i> score		
<i>M</i>	-2.95	-11.45
<i>SD</i>	25.52	27.96
<i>P</i> score adjusted for prethreat judgment of masculinity of partner		
<i>M</i>	+4.65	-4.76
<i>SD</i>	24.28	23.99

first at the gross differences between the experimental conditions. As shown in Table 2, the mean *P* score for the Favorable condition was -2.95; the mean for the Unfavorable condition was -11.45. Thus, subjects in the high dissonance condition tended to say that their partner's arousal level was about the same as their own, while those in the low dissonance condition tended to say their partner's arousal was somewhat less than their own. Attributing one's own characteristic to others was therefore more frequent in the high dissonance, or Favorable, group. The difference between the means yielded a *t* (for correlated means, due to matching) of 1.52, which is at the .13 level of significance. In order to arrive at a more firm conclusion regarding the outcome of variance which let us look at another source of variance which can be taken out of the gross variance in the *P* scores.

It had been anticipated that part of the variance in *P* scores would be due to the impression of masculinity created by the partner prior to the introduction of the undesirable cognition. If the subject rated his partner as very masculine on the masculine-feminine scale, then he would be likely to make somewhat lower (less homosexual) needle estimates for his partner than if he had rated him as very feminine. Correlations were therefore calculated between perceived masculinity of the partner (prethreat) and attribution of homosexual arousal to the photographs (*P* score). Within the Favorable group the resulting product-moment correlation was -.32 ( $p < .05$ ); in the Unfavorable group the correlation was -.52 ( $p < .01$ ). It will be remembered that the two groups did not differ in their mean (prethreat) rating of the



masculinity of the partner (as measured by the masculine-feminine scale); in addition, the distributions of these ratings were very similar in the two groups. Therefore, it was decided that the rather similar within-groups correlations would justify combining the groups, calculating the correlation between the two variables in the total sample, and computing adjusted  $P$  scores as deviations from the regression line. By means of this procedure the variance associated with how masculine the partner appeared (*prior* to the threat) could be partialled out. Within the total sample ( $N = 84$ ) the correlation between "masculinity" (prethreat) and  $P$  score was  $-.42$ . Each subject's  $P$  score deviation from the regression line was calculated, and the resulting scores were then interpreted as reflecting differences in attribution due to factors other than the initial perceived masculinity of the partner. The adjusted means for the groups are shown in Table 2. A  $t$  test for correlated means yielded a  $t$  of 2.04 ( $df = 40$ ), significant beyond the .05 level. On the basis of these results one may conclude that the groups differed in attribution in the direction predicted by the hypothesis relating dissonance and projection.

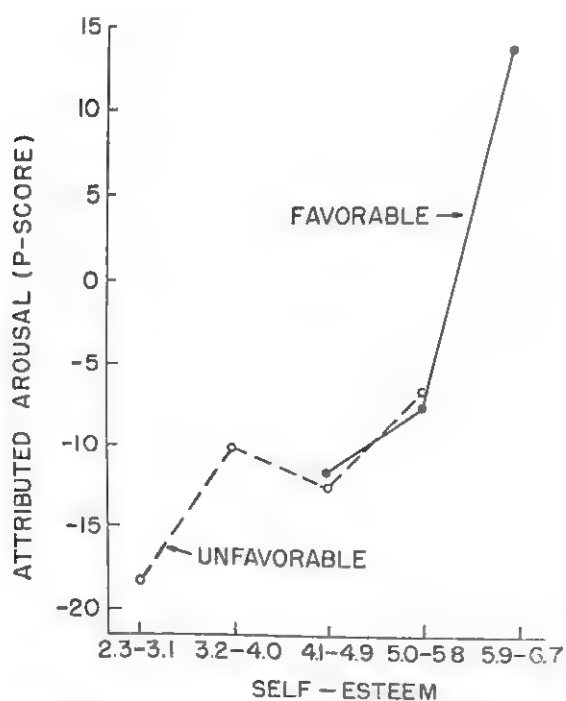


FIG. 1. Mean attribution of homosexuality as a function of level of self-esteem.

In all subsequent comparisons, the original, unadjusted  $P$  scores were used. The use of the simpler score should make interpretations clearer and more direct, especially in the case of within-conditions analyses.

It is of interest to look at the relation between self-esteem and attribution of arousal within the two experimental conditions. The product-moment correlation within the Favorable group was  $+.29$  ( $p < .07$ ); within the Unfavorable group the correlation was  $+.10$  ( $p < .55$ ). The relations are presented graphically in Figure 1, showing the mean raw  $P$  scores within each condition as a function of increasing self-esteem. The lack of correlation in the Unfavorable condition suggests that projection as a means of reducing dissonance occurred only when the amount of dissonance was quite high. From the point of view of the theory, this is not surprising. For the person with low self-esteem the undesirable information is actually consonant with his general self-evaluation (although dissonant with his specific cognitions about his adequate masculinity). Only for the person who believes he is consistently good will the undesirable information be strongly dissonant with his general self-esteem.

In Figure 1 it can be seen that the two groups show considerable continuity where they overlap in level of measured self-esteem. This fact is important because it implies that the self-esteem manipulations did not have strong opposed artifactual effects upon the amount of attribution of homosexual arousal. If the personality test report interviews had had effects on attribution in ways other than through the self-esteem variable, then differences between conditions might have appeared when considering subjects in the two groups with equivalent levels of self-esteem. Judging from Figure 1, persons in the two groups who had equivalent measured self-esteem levels apparently reacted similarly to the undesirable cognition.

*Projection and attitude toward available social objects.* There is a well known judgmental tendency which leads a person to perceive others as possessing traits consistent with his general evaluation of those others (a halo effect). On the basis of the halo effect alone, one would expect a tendency to at-

tribute homosexuality (an unfavorable trait) to persons who are evaluated in general relatively negatively. However, the presence of dissonance resulting from self-ascription of homosexuality should introduce a contrary tendency. To the extent that projection, of the type defined in this report, occurs, it should be aimed primarily at persons who are relatively favorably evaluated. Since in this experiment projection was expected to occur to a greater extent in the Favorable group, it follows that the empirical pattern of attribution in that condition should be some compromise between the projection pattern and the halo pattern, since one may expect both forces to be operating. In the Unfavorable condition, on the other hand, one would expect to find a pattern more closely resembling the pure halo pattern, due to the absence of large amounts of dissonance.

In Figure 2 the results are shown separately for the two conditions. Mean  $P$  scores are shown as a function of increasingly favorable evaluation of the partner, as measured independently and prior to the introduction of the cognitions regarding homosexuality. It can be seen that the results are consistent with the hypothesis. For relatively negative and moderate levels of evaluation of the partner, subjects in the Favorable and Unfavorable conditions attributed homosexual arousal consistent with a halo effect. The less favorably they rated the other subject, the more homosexuality they attributed to him. However, when the partner had been evaluated very favorably, subjects in the two experimental conditions reacted in quite different ways. The Unfavorable group continued to follow the halo pattern, attributing very low homosexuality to the partner. The Favorable group, in contrast, exhibited no decrease in attribution when confronted with favorably evaluated objects. In fact, there was a slight but insignificant increase. The difference between the mean  $P$  scores of the Favorable ( $n = 9$ ) and Unfavorable ( $n = 12$ ) groups at the high respect point (5.4 to 6.4 in Figure 2) was significant beyond the .05 level by the  $t$  test. Since the subject's respect for his partner was not experimentally manipulated, it is possible that via self-selection other variables may be contributing to the observed difference. It

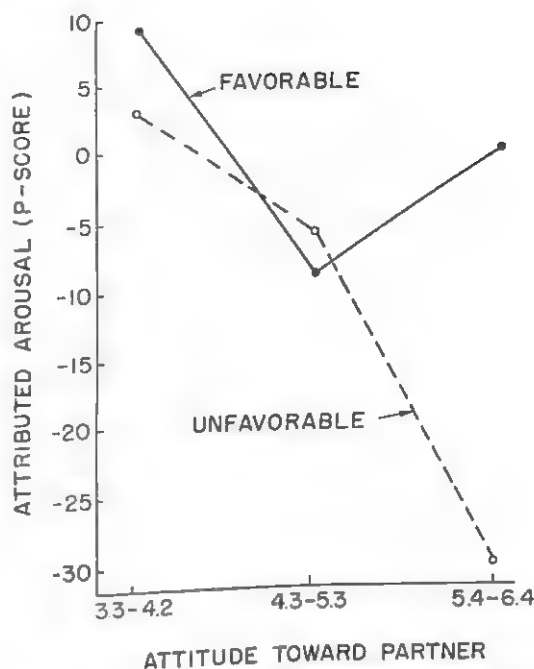


FIG. 2. Mean attribution of homosexuality as a function of favorability of attitude toward partner.

should be pointed out, for example, that the measured self-esteem level of subjects in the Favorable group (with respected partner) was slightly higher than that of other subjects in the Favorable group, so that some of the tendency of these particular subjects to project may be traceable to their higher self-esteem rather than to their attitude toward their partner. Unfortunately, the number of cases is too small to allow an internal analysis to throw light on this question. All things considered, one can be fairly confident that the difference between the conditions does reflect a tendency for projection to be directed toward favorably evaluated persons under these circumstances.

#### DISCUSSION

The results provided good support for the central hypothesis, that projection can be a response to dissonance involving the self-concept. Subjects in the Favorable condition, for whom the undesirable information about homosexuality was more dissonant with the self-concept, attributed more arousal to other persons. It is very unlikely that this difference was due to differences between conditions in severity of superego standards concerning homosexuality. In designing the ex-

perimental manipulations, care was taken to avoid any implications about the good or bad aspects of homosexuality. Since subjects were assigned randomly to conditions, it appears safe to assume that the groups did not differ on the average in their moral evaluation of homosexuality as such. The experiment was therefore capable of demonstrating the role of dissonance in projection while controlling the superego variable.

The finding that projection resulting from dissonance was aimed primarily at respected persons supported the second hypothesis. A number of defensive projection processes have been summoned in order to explain phenomena of prejudice toward out-groups (for example, Ackerman & Jahoda, 1950; Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). It is important to note that in most cases projection is said to be aimed at persons and groups who are disliked and considered incomparable and inferior to the projector. This is, of course, quite different from the kind of projection revealed in the present experiment.

Let us consider possible alternative explanations for the results shown in Figure 2. Perhaps the dissonance introduced by the undesirable information in the Favorable condition led directly to a re-evaluation of homosexuality. That is, perhaps these subjects were able to change their attitude toward the trait without the intermediate step of associating it with favorably evaluated persons. This re-evaluation prior to attribution might then affect the pattern of attribution in such a way as to give the appearance, deceptively, of projection. Once the undesirability of the trait was reduced, there would be less tendency to attribute it differentially to disliked persons (halo effect). The effect of such a process would be to attenuate the halo pattern in the Favorable condition, and might be revealed in part as a greater tendency to attribute homosexuality to respected persons, as compared with subjects in the Unfavorable condition. Is this hypothesis capable of explaining the results of the experiment without resort to the hypothesis of defensive projection?

The data show that subjects in the Favorable condition did tend to follow the halo

pattern when the partner was evaluated unfavorably or moderate in favorability. In Figure 2, the difference between the unfavorable ( $n = 9$ ) and moderate ( $n = 24$ ) points within the Favorable group was significant beyond the .10 level by the  $t$  test, and was at least as striking as the pattern for the Unfavorable group. If in fact the high dissonance in the Favorable condition had led directly to re-evaluation of homosexuality (without projection), then the halo effect would have been attenuated in this condition at all points in Figure 2. There is no apparent reason for supposing that re-evaluation of the trait (without defensive projection) would have occurred among Favorable subjects confronted with respected partners more so than among other subjects in the Favorable group.

Consider another alternative explanation of the results. Perhaps the personality test report led subjects in the Favorable condition to conclude that they were generally better than other people, while those in the Unfavorable condition concluded they were worse than other people. Both groups of subjects then were given information that they possessed some degree of homosexual arousal. If subjects in the Favorable condition believed they were generally superior people, they might then have deduced that their partner was likely to be less worthy than themselves. This could result in rating the partner as possessing a greater amount of homosexual motivation. Persons in the Unfavorable condition would, by similar reasoning, conclude that their partner possessed a smaller amount of homosexual arousal. Such nondefensive processes could account for the over-all difference in attribution between the two experimental conditions.

The data, as presented in Figure 2, cast doubt upon this alternative explanation. If the over-all difference between conditions were due to nondefensive deductions from the personality test reports, one would expect to find differences between the Favorable and Unfavorable conditions at all points along the attitude-toward-partner dimension. It is apparent, on the contrary, that the difference between conditions occurred only when the partner had been rated favorably.



It is interesting to speculate about the conditions under which dissonance with the self-concept will lead to projection onto favorably evaluated persons. In this experiment the dissonance producing information was probably quite striking and unambiguous to the subjects. With great care the experimenter had explained that movement of the needle in response to looking at the photographs was a clear and indisputable sign of homosexual arousal. The situation was such that outright denial of the meaning of the needle movements would have been quite difficult for persons in reasonable touch with reality. It is very likely that these subjects were forced to accept the information as implying some degree of homosexual arousal in themselves. Under these circumstances of self-ascription, a good way to reduce the dissonance remaining was to try to get desirable people "into the same boat."

However, if the subjects had been able to deny the direct implications of the galvanic skin response, then a different pattern of attribution might have been observed. If the information were sufficiently ambiguous, so that partial denial occurred, then it would no longer be so comforting to attribute the undesirable trait to persons with whom the subject ordinarily classes himself. That is, when one is attempting to avoid self-ascription, it probably does not help to ascribe the trait to others who are seen as generally similar to one's self. Whether, when denial is possible, projection tends to be directed toward undesirable persons or out-groups, is an interesting question for further experimental exploration.

#### SUMMARY

A laboratory experiment was conducted to investigate some of the conditions affecting the occurrence of defensive projection. It was hypothesized that such projection is a positive function of the amount of cognitive dissonance resulting from the introduction of a self-referent cognition of negative valence. Further, it was hypothesized that certain types of defensive projection are likely to be selectively aimed at persons who are favorably evaluated by the threatened individual. Two groups of normal subjects were prepared

in such a way that different amounts of dissonance would result from their exposure to the same undesirable information about themselves. All subjects received fraudulent information to the effect that they possessed homosexual tendencies. During the presentation of the disturbing material, each subject made estimates of the degree of homosexual arousal of another subject with whom he was paired, and whom he had met only rather briefly just prior to this part of the experiment.

The results supported the hypotheses. On the average, subjects in the high dissonance condition attributed to their partner about the same degree of arousal as they themselves appeared to be having. Those in the low dissonance condition in general attributed to their partner a level of arousal less than their own. The evidence suggested that the high dissonance group projected only when confronted with a partner whom they had previously evaluated quite favorably on adjective rating scales.

The relation between the psychoanalytic and dissonance theory approaches to defensive processes was discussed. It was proposed that study of the selection of objects should throw light upon the possible existence of distinct varieties of defensive projection.

#### REFERENCES

- ACKERMAN, N. W., & JAHODA, MARIE. *Anti-Semitism and emotional disorder*. New York: Harper, 1950.
- ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, 7, 117-140.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.
- GERARD, H. B. A specific study: Conflict and conformity. *Amer. Psychologist*, 1959, 14, 413. (Abstract)
- GERARD, H. B., & RABBIE, J. M. Fear, affiliation, and social comparison. *Amer. Psychologist*, 1960, 15, 409. (Abstract)
- MURSTEIN, B. I. The projection of hostility on the Rorschach and as a result of ego-threat. *J. proj. Tech.*, 1956, 20, 418-428.
- WRIGHT, BEATRICE A. Altruism in children and the perceived conduct of others. *J. abnorm. soc. Psychol.*, 1942, 37, 218-233.

(Received January 25, 1961)

## THE INFLUENCE OF HUNGER ON THE LEARNING AND RECALL OF FOOD RELATED WORDS<sup>1</sup>

SEYMOUR EPSTEIN

*University of Massachusetts*

AND

HERBERT LEVITT

*University Hospitals, University of Wisconsin*

Despite a number of studies on the steering or directive effects of primary drives<sup>2</sup> on perception, association, and imagination (see, for example, Clark, 1952; Epstein & Smith, 1957; McClelland & Atkinson, 1948; Murray, 1959; Wispé, 1954), not a single such study has been reported on learning in humans. As for retention, there are a few studies with humans on the directive influence of acquired drives (Rapaport, 1950), but none on primary drives.

The present study was undertaken to provide answers to the following questions: Does hunger at time of learning facilitate the learning and retention of words referring to food? Does hunger at time of recall facilitate the recall of words referring to food? Is there an interaction of hunger at time of learning and at time of recall upon recall of words referring to food? Are directive effects of hunger on learning and retention fairly general or do they vary as a function of specific characteristics of the stimulus?

A problem in working with the hunger drive in the laboratory is that if relatively high states of drive are to be obtained it is necessary to require abstinence from eating, which produces set effects that have been demonstrated to exert an influence as great as, or greater than, drive effects (Clarke & Epstein, 1957; Postman & Crutchfield, 1952; Taylor, 1956). Fortunately, several studies working within the normal food cycle have demon-

strated directive effects of drive as a function of 4-6 hours of deprivation (Epstein & Smith, 1956; Lazarus, Yousem, & Arenberg, 1953; Postman & Crutchfield, 1952; Sanford, 1936, 1937).

### METHOD

*Subjects.* Sixty subjects, paid at the rate of \$1.00 per hour, were enlisted from four fraternities at the University of Massachusetts. Subjects were randomly assigned to four equal groups according to a  $2 \times 2$  factorial design of hunger at time of learning and hunger at time of memory. Fifteen subjects were hungry at learning but not hungry at recall, 15 were hungry at both learning and recall, 15 were not hungry at learning but were at recall, and 15 were neither hungry at learning nor recall. Hungry subjects were obtained by testing at 5:00 P.M., before the evening meal, and control subjects by testing at 6:30 P.M., after the evening meal. The group that was not hungry at learning and hungry at recall had a 23-hour memory interval, whereas the group that was hungry at learning and not hungry at recall had a 25-hour memory interval. Although the difference of 2 hours might be expected to influence total recall, there was no reason to believe it would influence selective recall, and, as it turned out, total recall was unaffected.

*Procedure.* For the learning session, a paired-associates list containing food words and control words was presented in four different random orders by a modification of a group method described by Saltz and Myers (1955). Standard instructions for learning paired associates were given, and subjects were required to raise their hands after reaching the criterion of one trial without error for the entire list. In order to reduce temptation to misrepresent, an appeal to honesty was made, and subjects were told they would have to remain for the entire session no matter when they were finished. Moreover, as each trial was represented by a separate page, cheating would have required turning pages, which could have been detected by the experimenter. In order to prevent overlearning for the more rapid learners, subjects were required to work on anagrams task following completion of the learning task. From all observations it was evident that the instructions were adhered to. On the day following learning, three measures of memory were obtained. First the subject wrote down all the single words he recalled from the previous day, treating stimulus and response words independently. Next, the subject was given a test of recognition in

<sup>1</sup> This paper was presented, in part, at the American Psychological Association convention in 1958. The study was supported by Grant M-1293 from the National Institute of Mental Health, United States Public Health Service, as part of a project on the measurement of drive and conflict. At the time both authors were at the University of Massachusetts.

<sup>2</sup> Drive throughout this paper is defined as a state of activation with directive properties. For those who prefer to define drive as a state of activation only, the word "motive" could be substituted wherever the word drive appears.

which correct and incorrect words were included. Finally, a measure of relearning was obtained. The last two measures were discarded when it was found that scores were uniformly so high that there was not sufficient variance left to investigate.

**Paired-associates list.** The words to be learned consisted of 12 paired associates subdivided according to food relatedness and associative strength. High associative (HA) pairs consisted of a stimulus member paired with the word most often given to it in a preliminary word association test administered to 100 subjects from a comparable population. Low associative strength (LA) pairs consisted of a stimulus member paired with a word which it had never elicited as a response. There were four paired associates in which both members were food words—FRUIT-APPLE (HA), CHIESE-CRACKER (HA), WAFFLE-STEAK (LA), and CAKE-HAM (LA); four pairs in which both members were control words—CARPET-RUG (HA), LAMP-LIGHT (HA), ROOM-SOFA (LA), BED-WINDOW (LA); and four pairs in which one of the members was a food word and the other a control word, with order counterbalanced—EGG-STAIR (LA), STEW-CYLING (LA), HOUSE-BUTTER (LA), and SCREEN-POTATO (LA). The control words were related to home objects in order to partly control for the uniform concept formed by the food words. The pairs that contained words from both concepts, henceforth referred to as the "mixed" category, were necessarily of low associative strength, resulting in an overall design that was not balanced, but from which balanced designs could be extracted.

## RESULTS

**Learning.** Scores consisted of the first trial at which two successive correct anticipations occurred. The data were first treated by an analysis of variance in a three-dimensional design of hunger at time of learning, food relevance of paired associates, and associative strength of paired associates, each represented at two levels. The four paired associates involving mixed concepts were not included in this analysis as they did not occur at the level of high associative strength. In Table 1 it can be seen that hunger is significant as a main effect (.05 level). The hungry group learned words across classifications in fewer trials than the control group. However, there is no support for differential learning of pairs of food words and pairs of control words as a function of hunger (see sources of variance  $F \times H$  and  $A \times F \times H$ ). The remaining sources of variance, which are of only incidental interest, show that pairs of high associative strength were learned more rapidly than pairs of low associative strength (.01

TABLE 1

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION AS A FUNCTION OF HUNGER, ASSOCIATIVE STRENGTH, AND FOOD RELEVANCE

Source of variance	df	MS	F
Total between	59		
Hunger (H)	1	36.85	4.50*
Individual subjects (Ss)	58	8.19 <sup>a</sup>	
Total within	420		
Associative strength (A)	1	531.30	116.77**
A $\times$ H	1	16.50	3.63
A $\times$ Ss	58	4.55 <sup>a</sup>	
Food relevance (F)	1	107.35	41.61**
F $\times$ H	1	.61	—
F $\times$ Ss	58	2.58 <sup>a</sup>	
A $\times$ F	1	55.36	21.13**
A $\times$ F $\times$ H	1	2.84	1.08
A $\times$ F $\times$ Ss	58	2.62 <sup>a</sup>	
Words within categories (W)	4	23.14	8.67**
W $\times$ H	4	4.93	1.85
W $\times$ Ss	232	2.67 <sup>a</sup>	
Total	479		

\* Error term for the mean squares listed above it up to the next error term.

<sup>a</sup> Significant at .05 level.

\*\* Significant at .01 level.

level), and pairs of food words were learned more rapidly than pairs of control words (.01 level). There are significant differences (.01 level) between individual words within

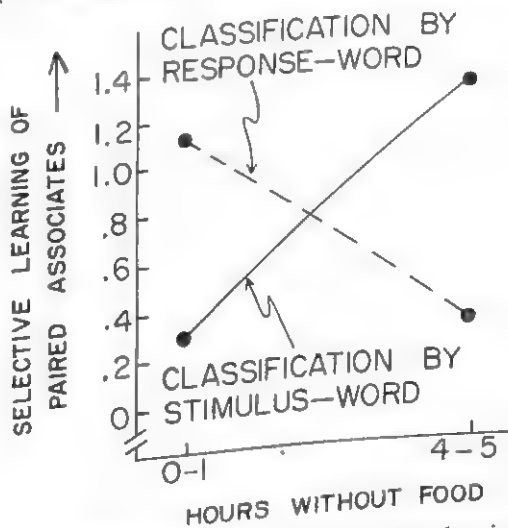


FIG. 1. Selective learning of food related paired associates as a function of whether the food word is a stimulus or response member. (Scores consist of mean trials to criterion for nonfood paired associates minus food paired associates, with food relevance determined independently for stimulus and response members. A high score indicates that food relevant paired associates were learned relatively rapidly.)



categories, and there is a significant interaction (.01 level) between associative strength and food relevance, with a greater disparity between the learning of pairs of food and control words of low associative strength than between the learning of food and control words of high associative strength.

In a second analysis of the learning data, the mixed pairs, which had not been previously included, were substituted for the pairs of high associative strength, thereby producing an orthogonal arrangement according to whether a word was a stimulus member or a response member and a food word or a control word. Thus, the following combinations of words were represented in the paired associates: food-food, food-control, control-food, and control-control. The analysis of variance was identical to the preceding analysis except that the classification by stimulus and response membership replaced the classification by low and high associative strength. It was found that the hungry group learned paired associates relatively more rapidly than the control group when the stimulus member was a food word and less rapidly when the response member was a food word. This can be seen in Figure 1 where mean trials to criterion for control words minus food words

TABLE 2

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION  
AS A FUNCTION OF HUNGER AND FOOD  
RELEVANCE OF STIMULUS AND  
RESPONSE MEMBERS

Source of variance	df	MS	F
Total between	59		
Hunger (H)	1	76.40	2.56
Individual subjects (Ss)	58	29.81	
Total within	420		
Stimulus relevance ( $S_r$ )	1	76.40	12.67**
$H \times S_r$	1	32.68	5.41*
$S_s \times S_r$	58	6.03 <sup>a</sup>	1.29
Response relevance ( $R_r$ )	1	57.86	14.57**
$H \times R_r$	1	18.48	4.65*
$S_s \times R_r$	58	3.97 <sup>a</sup>	
$S_r \times R_r$	1	129.64	21.72**
$H \times S_r \times R_r$	1	1.40	—
$S_s \times S_r \times R_r$	58	5.97 <sup>a</sup>	1.28
Individual words (W)	4	39.41	33.83
$H \times W$	4	23.73	5.09**
$S_s \times W$	232	4.66 <sup>a</sup>	
Total	479		

\* Error term for the mean squares listed above it up to the next error term.

\* Significant at .05 level.

\*\* Significant at .01 level.

TABLE 3

ANALYSIS OF VARIANCE OF NUMBER OF WORDS  
RECALLED AS A FUNCTION OF HUNGER AT  
TIME OF LEARNING, HUNGER AT TIME  
OF RECALL, ASSOCIATIVE STRENGTH,  
AND FOOD RELEVANCE

Source of variance	df	MS	F
Total between	55		
Hunger at learning ( $H_l$ )	1	.76	—
Hunger at recall ( $H_r$ )	1	.10	—
$H_l \times H_r$	1	1.45	1.02
Individual subjects (Ss)	52	1.41 <sup>a</sup>	
Total within	280		
Associative strength (A)	2	16.75	12.50**
$A \times H_l$	2	.08	—
$A \times H_r$	2	.25	—
$A \times H_l \times H_r$	2	3.18	2.37
$A \times S_s$	104	1.34 <sup>a</sup>	
Food relevance (F)	1	.29	—
$F \times H_l$	1	.30	—
$F \times H_r$	1	2.35	2.67
$F \times H_l \times H_r$	1	8.03	9.12**
$F \times S_s$	52	.88 <sup>a</sup>	
$A \times F$	2	11.76	8.98**
$A \times F \times H_l$	2	4.15	3.17*
$A \times F \times H_r$	2	2.36	1.80
$A \times F \times H_l \times H_r$	2	.95	—
$A \times F \times S_s$	104	1.31 <sup>a</sup>	
Total	335		

\* Error term for the mean squares listed above it up to the next error term.

\* Significant at .05 level.

\*\* Significant at .01 level.

is plotted as a function of time without food. (By subtracting food words from control words, rather than the other way around, all values were kept positive, and a high score made to indicate relatively rapid learning of paired associates containing food words.) The mean differences for the hungry and control groups are, respectively, 0.29 and 1.37 for classification by stimulus word and 1.12 and 0.31 for classification by response word. The analysis of variance indicates that both effects are significant at the .05 level (see Table 2, sources of variance  $H \times S_r$  and  $H \times R_r$ ). However the significant interaction of hunger and words within categories (see source of variance  $H \times W$ ) indicates that the findings may be a result of the particular words selected, and that verification with a different set of paired associates is required.

**Recall.** Recall was analyzed by an analysis of variance for the variables of hunger at learning, hunger at recall, food relevance of the word recalled, and associative strength, with the mixed category treated as an additional level of associative strength. Scores

were the number of words recalled within each of the six classifications formed by the two levels of food relevance and the three levels of associative strength. There were 14 entries of associative strength. There were 14 entries rather than 16 for each condition, as 2 subjects failed to appear for the recall session.

In Table 3 it can be seen that there is a significant interaction (.01 level) for the variables of hunger at learning, hunger at recall, and food relevance of the word recalled. In Figure 2 it is apparent that the reason for the interaction is that an increase in recall of food words as a function of hunger at recall occurs only when the groups not hungry at learning are compared. If only subjects not hungry at learning are considered, the mean number of food words minus control words recalled is  $-1.79$  for the condition of not hungry at recall, and  $0.93$  for the condition of hungry at recall. In contrast, when the same comparison is made for subjects who learned when hungry, the relative recall of food words varies little as a function of hunger at recall, with mean differences, respectively, of  $0.43$  and  $-0.29$  for the not hungry and hungry conditions. It is of interest that neither hunger at learning nor at recall approaches significance for total words recalled. Apparently, hunger decreases recall of control words to about the same extent that it increases recall of food words.

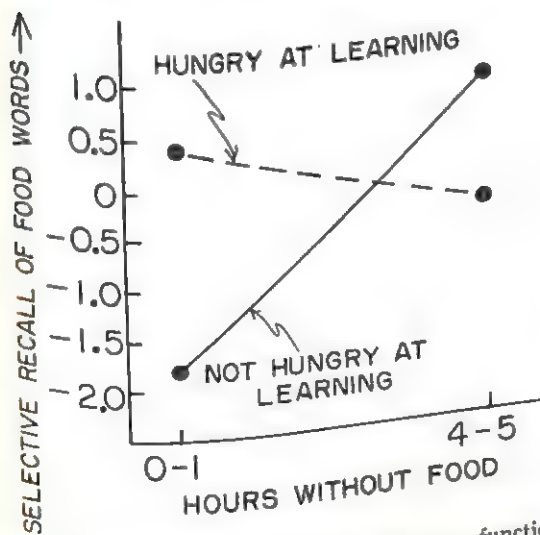


FIG. 2. Selective recall of food words as a function of hunger state during learning and during recall. (Scores consist of number of food words minus nonfood words recalled.)

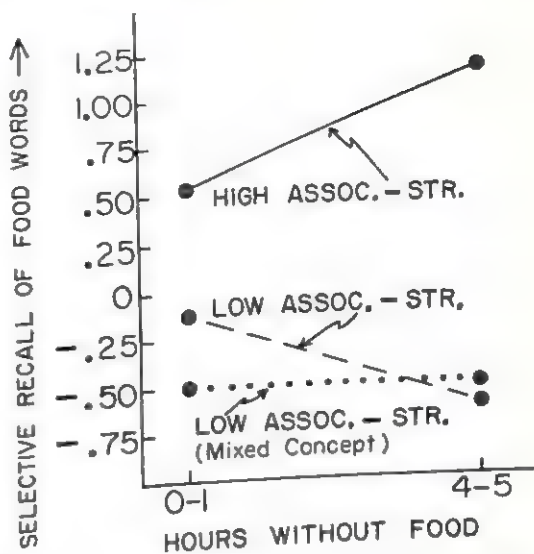


FIG. 3. Selective recall of food words as a function of hunger at learning and associative strength of the paired associates presented in the learning session. (Scores consist of number of food words minus nonfood words recalled.)

There is a significant interaction for the variables of hunger at learning, food relevance, and associative strength (.05 level). In Figure 3 it can be seen that the relationship is a direct one for hunger at time of learning and relative recall of food words for the food words which had previously been presented in high associative strength pairs, and an inverse one for the food words presented in low associative strength pairs. No relationship between relative recall of food words and hunger at recall is indicated for food words of low associative strength presented in mixed concepts.

#### DISCUSSION

It would seem reasonable to expect that hunger at time of learning would facilitate the learning of food words, that hunger at time of recall would facilitate the recall of food words, and hunger at both times would result in the greatest degree of selective recall. Brown (1953), in a highly relevant discussion of the influence of hunger upon perception states that,

it would not be surprising if the hungry Ss, who possess both the internal cues and the visual cues, should do better than the nonhungry Ss, who

posses only the visual cues. On the present view, the hungry group might also be superior to the non-hungry group on nonfood pictures, though their superiority should be less than in the case of the food pictures. Facilitation of performance on non-food pictures would be attributed to the energizing effects of drive *per se*, not to the presence of hunger stimuli (p. 9).

Assuming that the same principles apply to learning as to perception, the finding of a general facilitating effect of hunger upon learning supports the position illustrated by Brown on the significance of the activating role of drive. However, the results on selective learning and recall appear to be far more complex than would be the case if drive produced cues simply combined with stimulus produced cues to facilitate drive related responses. Several studies on the influence of hunger upon perception and imagination have suggested that associated with every drive state there is a drive oriented autistic process and a reality oriented inhibitory process (Epstein, 1961; Epstein & Smith, 1956; Levine, Chein, & Murphy, 1942; McClelland & Atkinson, 1948; Sanford, 1937; Wispé, 1954). It is possible that the finding of facilitation in learning when the stimulus member is a food word was an example of the autistic process, and that the retardation in learning when the response member was a food word of the inhibitory process. If this is so, it would suggest that drives tend to induce sensitization in perception by inducing set effects or by some other means, and to induce defensiveness in overt response.

In regard to recall, it was found that the directive effect of hunger upon recall depended upon the hunger condition at learning. Only when learning took place in the nonhungry condition was selective recall as a function of hunger at recall demonstrated. This would suggest that drive produced cues can interfere with as well as facilitate recall of drive related material.

The finding that hunger increased recall of food words without influencing total recall suggests that the directive effects of a drive may function as an attention or set producing phenomenon, i.e., by focusing attention in one direction, attention in other areas is diminished. In this respect, drive produced set

effects differ from other set effects only to the extent that the cues are internal and therefore less likely to be correctly labeled.

The effect of hunger on selective learning was more equivocal than its effect on selective recall. This is not surprising, as recall involves a greater separation from the stimulus than does learning. An increase in memory interval can be viewed in the same manner as an increase in ambiguity, i.e., both factors by reducing the degree to which the response is determined by the stimulus, increase the degree to which motivational and incidental factors play a role.

The finding that hunger during learning resulted in an increase in selective recall of food words of high associative strength but a decrease in recall of food words of low associative strength can be understood by considering the particular combinations of words investigated. High associative strength pairs consisted of combinations, such as CHEESE-CRACKER, which are commonly experienced together during reduction of hunger, while the same does not apply to the low associative strength combinations, such as WAFFLE-STEAK. It is likely that a drive arouses only those associations that have previously been associated with it or with its reduction, rather than all stimuli related to the drive in the individual's cognitive framework.

An obvious limitation in the present study is the very mild state of drive investigated. Coupled with the empirical nature of the study and the unanticipated findings, the need for verifications is apparent. However, it would seem that the "new look" is as appropriate for learning as it is for perception. Apart from theoretical considerations, the directive influence of drives on learning is of considerable practical significance and deserves more experimental attention than it has received.

#### SUMMARY

The study was undertaken to investigate the influence of hunger on the learning and recall of words related to food. Thirty subjects learned a list of paired associates consisting of all possible combinations of food and nonfood words, immediately before the



evening meal. Thirty other subjects learned the same paired associates immediately after the evening meal. Half of each group was tested for recall the next day before the evening meal, and half after the evening meal. The recall task consisted of writing all the single words that could be remembered from the learning session of the previous day.

The major findings and conclusions were as follows:

1. The group hungry at learning learned more rapidly than the control group.
2. The group hungry at learning demonstrated relative facilitation in the learning of paired associates when the stimulus word was a food word, and relative retardation when the response word was a food word. This was interpreted as indicating that drives tend to induce selective facilitation in perception but inhibition in response for drive related material.
3. There was an interaction of hunger condition at learning and at recall upon selective recall of food words. Only when comparison was between the groups who were not hungry during learning was hunger during recall found to be directly associated with selective recall of food words. It was concluded that drive produced cues interfere with as well as facilitate recall of drive related material.
4. Increased recall of food words as a function of hunger occurred at the expense of a decrease in recall of control words. It was suggested that drive creates a selective set for drive related materials and consequent inattention to other material.
5. The group hungry at learning exhibited greater selective recall of food words presented in high associative strength pairs and less selective recall of food words presented in low associative strength pairs than the group not hungry during learning. It was concluded that drive tends to activate associations of stimuli that have been present during arousal or reduction of the drive, rather than of all stimuli related to the drive in the individual's cognitive framework.

## REFERENCES

- BROWN, J. S. Problems presented by the concept of acquired drives. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1953*. Lincoln: Univer. Nebraska Press, 1953. Pp. 1-21.
- CLARK, R. A. The projective measurement of experimentally induced levels of sexual motivation. *J. exp. Psychol.*, 1952, 44, 391-399.
- CLARK, A. R., & EPSTEIN, S. Food-related responses to ambiguous stimuli as a function of time without food and experimental set. *Amer. Psychologist*, 1957, 12, 394. (Abstract)
- EPSTEIN, S. Food-related responses to ambiguous stimuli as a function of hunger and ego strength. *J. consult. Psychol.*, 1961, 25, 463-469.
- EPSTEIN, S., & SMITH, R. Thematic apperception as a measure of the hunger drive. *J. proj. Tech.*, 1956, 20, 373-384.
- EPSTEIN, S., & SMITH, R. Thematic apperception, Rorschach content, and ratings of sexual attractiveness of women as measures of the sex drive. *J. consult. Psychol.*, 1957, 21, 473-478.
- LAZARUS, R. S., YOUSEM, H., & ARENBERG, A. Hunger and perception. *J. Pers.* 1953, 21, 312-328.
- LEVINE, R., CHEIN, I., & MURPHY, G. The relation of the intensity of a need to the amount of perceptual distortion. *J. Psychol.*, 1942, 13, 283-293.
- MCCLELLAND, D. C., & ATKINSON, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, 25, 205-222.
- MURRAY, E. J. Conflict and repression during sleep deprivation. *J. abnorm. soc. Psychol.*, 1959, 59, 95-101.
- POSTMAN, L., & CRUTCHFIELD, R. S. The interaction of need, set, and stimulus-structure in a cognitive task. *Amer. J. Psychol.*, 1952, 65, 196-217.
- RAPAPORT, D. *Emotions and memory*. (2nd ed.) New York: International Univer. Press, 1950.
- SALTZ, E., & MYERS, T. I. A method for group presentation of paired-associate learning materials. Paper read at Midwestern Psychological Association convention, Chicago, 1955.
- SANFORD, R. N. The effect of abstinence from food upon imaginal processes: A preliminary experiment. *J. Psychol.*, 1936, 2, 129-136.
- SANFORD, R. N. The effect of abstinence from food upon imaginal processes: A further experiment. *J. Psychol.*, 1937, 3, 145-159.
- TAYLOR, JANET A. Physiological need, set, and visual duration threshold. *J. abnorm. soc. Psychol.*, 1956, 52, 96-99.
- WISPE, L. G. Physiological need, verbal frequency, and word association. *J. abnorm. soc. Psychol.*, 1954, 49, 229-234.

(Received February 2, 1961)

## VALUES EXPRESSED IN AMERICAN CHILDREN'S READERS: 1800-1950<sup>1</sup>

RICHARD DECHARMS AND GERALD H. MOELLER

Washington University

Students of cultural change within the United States seem to have reached some agreement as to a trend observable within the last century. This trend, which deals with some of the basic values of our culture, may be seen as a change from what Weber (1930) called "the Protestant ethic" to what has been called the "social ethic" (Whyte, 1956). Specifically, the dominant value of individual salvation through hard work, thrift, and competition is seen as being replaced by "a belief in the group as the source of creativity; a belief in 'belongingness' as the ultimate need of the individual; and a belief in the application of science to achieve the belongingness" (Whyte, 1956, p. 7). In Riesman's (Riesman, Glazer, & Denney, 1950) terminology the basic trend is from inner-direction to other-direction.

Actually this process is circular in the sense that the cultural change is probably accompanied by a change in values which starts a new cycle. The psychologist likes to conceive of the basic variables in human behavior as being internally determined and thus breaks this circle and concentrates on motives as basic.

The aim of the present paper is to investigate psychological variables which it seems logical to predict will be associated with the cultural changes observed in the United States over the last century and a half.

McClelland (1955) has noted striking similarity between his concept of the person with high achievement motivation and Riesman's inner-directed character type. McClelland defines achievement motivation as "success in competition with a standard of excellence"

(p. 43). According to Riesman et al. (1950) "the drive instilled in the child is to *live up to ideals* and to test his ability to be on his own by continuous experiments in self-mastery" (p. 59). For the other-directed person "making good becomes almost equivalent to making friends" (Riesman et al., 1950, p. 66). This sounds very much like the person with high affiliation motivation. Atkinson (1958) defines affiliation motivation as "concern . . . over establishing, maintaining, or restoring a positive affective relationship with another person. This relationship is most adequately described by the word friendship" (p. 206).

Assuming that achievement motivation is a basic component of the inner-directed character type, and that affiliation motivation is a basic component of the other-directed character type, in the context of Riesman's cultural change thesis, one would predict a decline in over-all achievement motivation and an increase in affiliation motivation in the last century of United States history. Strauss and Houghton (1960) have found evidence giving some support to these hypotheses in the period since 1924 in a study of 4-H club journals. A meaningful relationship should also be found between achievement orientation and economic and technological change according to McClelland (1955).

A further aspect of Riesman's thesis is that the stage of inner-direction is preceded by a stage of tradition-direction. During this stage strict moral codes demand behavioral conformity of the individual. The change from a tradition-directed society to an inner-directed society involves a secularization in the sense that the individual must prove himself worthy as in Weber's (1930) Protestant ethic, rather than be told what to do by categorical imperatives. One might thus predict more reliance on moral teaching early in the history of the United States.

<sup>1</sup> This study was carried out under Contract Nonr-816 (11) between the Office of Naval Research and the Small Groups Laboratory, Social Science Institute at Washington University.

We wish to thank David C. McClelland and Raymond E. Callahan for critical comments on an earlier version of the paper.



The objective measurement of cultural orientations or values is always difficult, especially if an attempt is to be made to tap the values of the past. An intriguing attempt to measure the motives of an ancient culture has been reported by McClelland (1958). He has developed a method of assessing achievement motivation in a culture by content analysis of literary products of the culture. Using this tool he found striking confirmation for his hypothesis that achievement motivation preceded the economic and technological development of Athenian civilization in classical Greece, a culture also discussed by Riesman.

McClelland's measure of motivation was developed originally to assess the motives of individuals (McClelland, Atkinson, Clark, & Lowell, 1953). The extension of it to apply to cultures raises questions as to what is being measured. McClelland (1958) has argued that his content analysis of documents produced by the culture will give a measure of the level of specific types of motivation within the culture if the documents are carefully selected to reduce effects of other obviously important variables. The best support for this argument derives from a study which demonstrated that content analysis of American Indian folk tales relates to content analysis of stories obtained from individuals in the manner used originally to validate the measure for studying individual motives (McClelland & Friedman, 1952).

It is obvious that a measure which has been shown to be related to individual motives would be expected to reflect the motives of the writer of any document chosen from a culture. In order to use such a measure as an indication of cultural orientation one does not have to assume that the motive score of an individual author is a measure of the cultural orientation alone. One must, however, assume that a portion of the score is a measure of the cultural orientation. The problem of reducing idiosyncratic components in the measure of cultural orientation becomes one of (a) sampling randomly many authors (b) under as similar conditions as possible (i.e., writing similar material) and (c) choosing materials which should place few restrictions on the author's fantasy.

All of the above advantages can be ob-

tained by careful sampling of stories written for children's readers. In addition, the stories are actually written to be used in transmitting cultural values, and information on how widely they have been used gives at once some indication of cultural acceptance of the values contained in the book and of the extent of its influence. An example of the use of children's readers to assess values in many cultures has been presented by McClelland (1961).

The schema presented above predicts a relationship between cultural achievement orientation and behavior of the members of the culture which would lead to technological advance. Just as one might predict that an individual with high achievement motivation might strive for some unique accomplishment, one might also predict that a culture with strong achievement orientation would produce many inventions. A measure of the inventiveness of the culture at various periods in history might be obtained from the number of patents issued per population, and one could predict a relationship between this and a measure of cultural achievement orientation.

### *Hypotheses*

The present study is an attempt to plot the incidence of achievement and affiliation imagery and moral teaching in a sample of children's readers from 1800 to the present. In addition, achievement imagery is to be related to data on the number of patents issued per population.

*Hypothesis I.* The incidence of achievement imagery in a sample of children's readers selected over the period 1800-1950 will decrease over the time period.

*Hypothesis II.* The incidence of affiliation imagery in the same sample of readers will increase over the time period.

*Hypothesis III.* The incidence of moral teaching in the sample will decrease over the time period.

*Hypothesis IV.* The incidence of achievement imagery will be positively related to the number of patents issued, corrected for changes in population.

The hypotheses assume, with Riesman, that the nineteenth century in the United States was dominated by the inner-directed charac-



ter type. Riesman is not specific as to dates, but it would appear that the early period of the century witnessed the transition from the tradition-directed character type and that the United States has recently been in transition from an inner-directed phase to an other-directed phase.

### METHOD

A bibliography of reading textbooks with copyright dates ranging from 1800 to 1952 was compiled. An attempt was made to procure at least four books from each 20-year period beginning in 1800. Readers were excluded which were not in wide use<sup>2</sup> during the period or which were used by religious affiliated schools. In the more recent periods from which more than four books were available, the choice of books was made randomly. In the periods in which fewer than four books were available the sample from each book was enlarged in so far as possible. Generally, the sample from each book was obtained by scoring every third page.

It was found that the number of words per page was sufficiently similar throughout the total sample to allow use of the page itself as the scoring unit. In order to equate for number of pages available the score was the number of pages containing imagery per 75 pages sampled. A raw score was thus computed for every 75 pages sampled (i.e., 25 pages scored).

The readers chosen for the study were, generally speaking, at a fourth grade level. During the nineteenth century many readers were designated in ways which had no relevance to grade level or, as in some cases, grade level was quite different from that of contemporary American readers in which the vocabulary is based on standard word lists. In some instances it was necessary to use the Dale and Chall (1948) formula<sup>3</sup> for predicting readability to determine whether the readers might be allowed in the study.

The pages selected from each book were scored

<sup>2</sup> Evidence of "wide use" was fairly easy to establish in readers published after 1850 since the McGuffey readers and the readers of the major book companies enjoyed national popularity as official texts of large school systems. Prior to 1850 Johnson (1904) and Nietz (1961) provided lists of historical texts from which to choose. In the first two decades of the sample the only secular texts available were used.

The following criteria were used in this order to establish "wide use": Evidence (often cited in later editions) of number of copies sold; several editions of the same book; and knowledge of use by large school systems.

<sup>3</sup> This readability formula which utilizes a word list and sentence length for determining grade level of reading materials was used because of evidence of its applicability to middle grade reading matter and high correlations with other formulae and criteria of readability (Chall, 1958).

independently by two scorers as to whether the page contained (a) achievement imagery, (b) affiliation imagery, or (c) a category called moral teaching. Achievement and affiliation imagery were scored according to the procedure outlined in Atkinson (1958). The subcategories usually scored in this procedure were not scored. The category Moral Teaching was developed and defined as explicit or implicit statements of judgment between right and wrong from the point of view of the author. The following (McGuffey, 1857) are examples of items which were scored for moral teaching:

The little boy took care of his faithful dog as long as he lived and never forgot that we must do good to others, if we wish them to do the same to us (p. 42).

Now that is the way with a great many thoughtless, quick tempered people. They try to find fault with somebody or something else, and get into a passion, and perhaps do mischief, when, if they would but reflect a little, it is their own dear selves who ought to bear the blame (p. 47).

Scorer reliabilities, based on presence of imagery only, were consistently high (Achievement Imagery = 94%, Affiliation Imagery = 96%, Moral Teaching = 97%).

The number of patents issued by the United States Patent Office and the United States Census figures were taken from governmental documents (United States Department of Commerce, 1960) and a patent index was computed by dividing the number of patents granted in a 20-year period by the population reported in the midyear of that period and multiplying by one million. This results in an index of patents issued per one million population during the period.

There are two methodological flaws in the procedure which it was felt might have had an effect on the results. In the first place, the technique of blind scoring was not employed. The scoring was done directly from the book and it was therefore probable that the scorer knew the date of the book. The effect of this knowledge cannot be assessed. A second methodological flaw lies in the sampling procedure. Systematic samples were taken from each book and the books were chosen as representative and in wide use. However, since each score was based on 75 pages of text, some books were more heavily weighted than others and the individual values of their authors might have unduly influenced the results.

In order to correct these methodological flaws it was decided to repeat the study with a drastically smaller sample. A sample of 6 pages was chosen at random from each book. The sampling of books followed the same criteria as those used in the first sample. Four books were selected from each period except the periods 1800-1819 and 1820-1839 where only two were available. The books from these two periods were double sampled. Scores on each variable were assigned to each book giving four scores for each of the eight periods thus resolving the ambiguities of sampling in the first study.

The sample had to be drastically cut since the

pages were typed and coded for blind scoring. (A total of 192 pages were scored independently by two scorers in the replication whereas 2,375 pages were scored in the first study.) The same number of typed lines were taken starting with each page which had been drawn randomly.

Whenever available different books were selected for the second sample. It was anticipated that since this was a much smaller sample than the first the results would not be as statistically significant. It was felt that general trends in the same direction would validate the statistically significant findings of the first sample. Actually, plots of the results of the two studies are almost identical and statistical significance was reached in most instances in the second study, although, as anticipated, the probability levels were not as great as in the first. This comparison of the two replications gives greater confidence in the results of the first study. The data presented here come from the first study. Statistical analyses will be presented for both studies.

### RESULTS

Table 1 presents the mean imagery scores for achievement, affiliation, and moral teaching in each of the 20-year periods. Figures 1 and 2 are graphic presentations of these data.

Hypothesis I predicted a consistent decrease in achievement imagery. The data (see Table 1 and Figure 1) show a sharp decline since 1890, but a steady increase from 1800 to the peak at about 1890. The second sample showed almost an identical curve with consistent increase up to about 1890 and then a sharp decline. The data of both samples show a significant relationship between amount of imagery and date (First sample,  $F = 8.09$ ,  $df = 7, 87$ ,  $p < .0005$ ; Second sample,  $F = 2.62$ ,  $df = 7/24$ ,  $p < .05$ ).

TABLE 1

MEAN SCORES FOR ACHIEVEMENT IMAGERY, AFFILIATION IMAGERY, AND MORAL TEACHING IN THE FIRST SAMPLE OF READERS

20-year midpoint	N <sup>a</sup>	Achievement imagery		Affiliation imagery <sup>b</sup>		Moral teaching <sup>b</sup>	
		$\bar{X}$	$\sigma^2$	$\bar{X}$	$\sigma^2$	$\bar{X}$	$\sigma^2$
1810	3	2.67	4.5	3.33	2.5	16.00	22.0
1830	3	2.50	1.7	4.25	11.7	16.75	9.7
1850	4	2.50	1.7	4.25	11.7	16.75	11.2
1870	12	4.42	13.4	6.00	24.4	12.42	1.0
1890	3	8.33	2.5	6.33	25.5	6.00	4.9
1910	16	11.06	13.5	5.13	4.7	4.19	2.7
1930	10	9.40	5.1	6.70	12.0	4.50	0.7
1950	15	6.33	19.6	9.33	6.1	1.00	0.1
	32	4.25	14.8	5.50	12.7	0.06	

<sup>a</sup> Number of samples of 25 pages scored. The raw score was the number of pages (out of 25) containing imagery.

<sup>b</sup> Variance heterogeneous. In no case were the variances heterogeneous in the second sample.

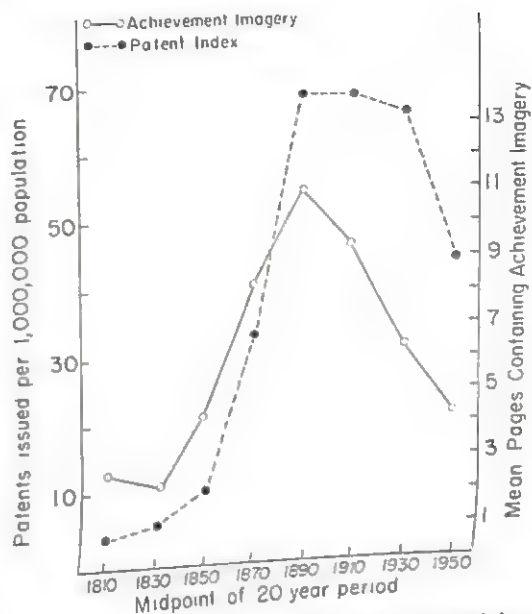


FIG. 1. Mean number of pages (out of 25) containing achievement imagery and the patent index.

Hypothesis II predicted a consistent increase in affiliation imagery from 1800 to the present. The data in Table 1 and Figure 2 show no consistent increase but do show a general trend with an unexpected drop in 1950 to the 1890 level. Analysis of variance shows a significant relationship ( $F = 2.41$ ,  $df = 7, 87$ ,  $p < .05$ ) but the variance was heterogeneous. The Kruskal-Wallis (see Siegal, 1956) analysis of variance of ranks also shows a significant relationship ( $p < .01$ ). In addition, a rank correlation (Kendall's tau) between affiliation imagery and date equals  $+ .61$  ( $p < .03$ ).

The data from the second sample show the same general trend with greater variability. The correlation drops below significance and analysis of variance does not reach significance.

Hypothesis III predicts a decrease in moral teaching from 1800 to the present. The data show a striking confirmation (First sample,  $F = 101.9$ ,  $df = 7/87$ ,  $p < .0005$ , heterogeneous variance; Second sample,  $F = 6.95$ ,  $df = 7/24$ ,  $p < .0005$ , variance homogeneous).

Hypothesis IV predicts a relationship between the amount of achievement imagery during a specific period and the index of patents issued in graphic form. Figure 1 shows this relationship in graphic form. A striking relationship is apparent. The rank correlation

(tau) here is + .79 ( $p < .003$ ). The correlation in the second sample is + .68 ( $p < .02$ ).

DISCUSSION

*Achievement imagery and patent index.* The data on achievement imagery do not confirm the original hypothesis which was obviously too simple. There are clear indications in both samples that achievement imagery increases to a peak around the turn of the present century and has steadily declined since then. This relationship is supported by the strikingly similar data from the patent index. The number of patents granted was used in preference to the number of applications for patents for two reasons: no record was kept of patents applied for until 1840 (this would have cut 39 years from the patent/population measure) and the very fact that a patent is issued is indicative of the "uniqueness" of the patent. Unique accomplishment is one of the criteria for scoring achievement imagery (Atkinson, 1958).

*Affiliation imagery.* The data from both samples tend to confirm the hypothesis of increasing affiliation imagery, although the results were not statistically significant in

the second sample. Certain aspects of the difference in type of affiliation imagery through time are noteworthy. Much of the early and middle nineteenth century readers' affiliation imagery, though widely scattered, was quite unsophisticated as shown by the following (Parker & Watson, 1857) example:

I love my dear little brother and I am pleased when I see him happy. I did not intend to disobey you, dear father, and I hope you will not be displeased with me for what I have done (p. 75).

The discussion of the joys of giving to others continued unabated for the next page and a half. In contrast, affiliation imagery in the period from 1920 was considerably more subtle.

It was more difficult to score affiliation imagery in the earliest books in the study and it was in this period that interrater reliability was lowest. The difficulty lay in differentiating true affective affiliation imagery from a culturally sanctioned form of address. Thus "dear son," "dear father," and "my dear" were not scored unless affect was also demonstrated since this was often mere conventionally approved formalism.

It is possible that this scoring difficulty could account for the results found. This, in combination with the fact that the results were not significant in the second sample, suggests caution in interpretation of these results.

*Moral imagery.* The decline in the religious-moral emphasis in textbooks has long been noted by various researchers. Hart (1933), in analyzing selected popular magazines of the period from 1900 to 1930, found evidence of a general decline in the status of religion and religious sanctions. These findings are in general accord with the results of the present study. In the case of the school readers it may be argued that the diminishing frequency of moral references is a result of the secularization of the schools during the nineteenth century.

It should be noted that the first schools in the colonies were church-sponsored and, in many instances, the minister of the church also served as teacher to the children. His primary purpose was not general education but simply to teach the pupils to read the Bible. The shift from church-sponsored to the

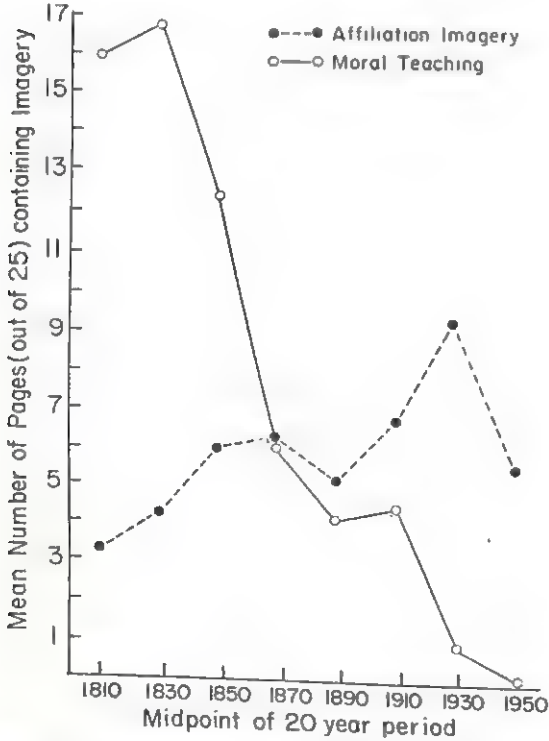


FIG. 2. Mean number of pages (out of 25) containing affiliation imagery and moral teaching imagery.



public-sponsored and supported schools began about the time of the American Revolution and continued through most of the nineteenth century. However, since all the books in the present study were prepared for public schools, the decrease in moral teaching imagery indicates the cultural trend toward secularization which affected the management of the schools and was reflected in the books written for the schools even after they had become nominally secular.

The antecedent conditions of changes in values such as demonstrated here are very complex. As noted earlier theorists such as Riesman et al. (1950) stress the importance of psychological factors (character type) which lead to the examination of child rearing practices. Economic historians stress political and economic factors (Rostow, 1960). The historical evidence is probably easier to marshal to support the economic interpretation.

The findings of the present study fit very well the conceptual paradigm of Rostow (1960) who stresses the economic factors. He has developed a general model of the stages of economic growth which distinguishes (a) precondition for take-off, (b) the take-off, (c) the drive to maturity, and (d) the age of high mass consumption.

In the United States, Rostow (1960) found that the traditional or agricultural society lasted until about 1840. The take-off occurred from 1843 to 1860. This appears to be the take-off period for achievement orientation also (see Figure 1). During the next period from 1860 to 1900, called by Rostow the drive to maturity, "some ten to twenty per cent of the national income was steadily invested, permitting industrial output regularly to outstrip the increase in population" (p. 9). The United States, according to Rostow's data and reasoning, reached technological maturity around 1900. This date is extremely close to the high points of achievement imagery and patent measures.

Rostow's (1960) preconditions for take-off are technological developments which might set the stage for increased social mobility, a factor mentioned by Riesman et al. (1950) as affecting child rearing practices. Rostow feels that during this period the idea that economic progress is possible and necessary for such ends as national dignity becomes

prevalent in the culture and men come forward who are willing to mobilize savings, take risks, and engage in entrepreneurial activity.

The latter aspects have a distinctly psychological flavor. Men who take risks and engage in entrepreneurial activity are those who have high achievement motivation (McClelland, 1958). Recently economists have noted the importance of motivation and personality structure in economic growth. Thus Hagen (1958) discusses the role of the need for achievement, for autonomy, for aggression, for dominance, for affiliation, and for dependence in the beginning of economic growth. These motivational variables interact with economic and political variables to produce cultural changes.

As noted in the introduction, evidence for the importance of psychological factors such as motives and values in cultural change and economic growth has been presented by McClelland (1955, 1958, 1961). There are, moreover, studies which have made a start in uncovering the relationship between child rearing practices and achievement and affiliation motivation. Briefly, achievement motivation appears to be associated with early parental stress on independence training and mastery, coupled with a warm acceptance of the child (Winterbottom, 1958). Affiliation motivation is related to maternal acceptance and to parental stress on interpersonal involvement of the child (Gall, 1960).

These findings appear to be in accord with Riesman's analysis of the child rearing practices which lead to the inner- and other-directed character types. The parent rearing a child in the period of transition to inner-direction must equip him with a "gyroscope" which will fit him to remain on course in a society where it is impossible to foretell, due to increasing social mobility, what role he will be called upon to play. He must be equipped to be self-reliant and independent. These are the aspects which Riesman sees in nineteenth century child rearing. The antecedents of achievement motivation seem clear.

On the other hand, with increasing urbanization and population density which result from technological advance, the child is no longer pushed to be independent, but learns the importance of other individuals in the

environment. He must be taught to win approval. Although Riesman's argument is more complicated than this, the child rearing practices which he sees in contemporary United States culture seem to be ones which might lead to affiliation motivation.

In summary we propose that motivation, or cultural orientation, be conceived of as an intervening variable standing between antecedent environmental factors associated with economic and political changes and consequent behavior resulting in cultural changes such as technological growth. Such an analysis should give increased explanatory power, since it is probable that motivation is a function of factors other than economic changes. For instance, cultural values affect child rearing practices and hence motives (McClelland, Rindlisbacher, & deCharms, 1955). Thus two cultures undergoing similar economic or political change may react quite differently due to the intervening variables of values, child rearing practices, and motives.

#### SUMMARY

Content analysis of children's readers from schools in the United States demonstrated a rise in achievement imagery from 1800 to about 1900 and then a steady decline. The achievement imagery curve was related to an index of the number of patents issued. A steady decline over the period 1800-1950 was found in the amount of moral teaching in readers. There was tentative indication of an increase in affiliation imagery.

The over-all picture presented by the data correspond very well with certain cultural trends pointed out by students of cultural change. The data illustrate an interesting technique for obtaining objective data to investigate cultural historical hypotheses.

#### REFERENCES

- ATKINSON, J. W. *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.
- CHALL, JEANNE. Readability: An appraisal of research and application. *Bur. educ. Res. Monogr. Ohio St. U.*, 1958, No. 34.
- DALE, E., & CHALL, JEANNE. A formula for predicting readability. *Educ. res. Bull.*, 1948, 27, 11-20, 37-54.
- GALL, H. S. The development of affiliation motivation. Unpublished doctoral dissertation, University of North Carolina, 1960.
- HAGEN, E. E. How economic growth begins: A general theory applied to Japan. *Publ. opin. Quart.*, 1958, 12, 373-390.
- HART, H. Changing social attitudes and interests. In *Recent social trends in the United States*. New York: McGraw-Hill, 1933. Pp. 382-442.
- JOHNSON, C. *Old-time schools and school-books*. New York: Macmillan, 1904.
- MCCLELLAND, D. C. Some social consequences of achievement motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1955*. Lincoln: Univer. Nebraska Press, 1955. Pp. 41-65.
- MCCLELLAND, D. C. The use of measures of human motivation in the study of society. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 518-552.
- MCCLELLAND, D. C. *The achieving society*. Princeton: Van Nostrand, 1961.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Croft, 1953.
- MCCLELLAND, D. C., & FRIEDMAN, G. A. A cross-cultural study of the relationship between child-rearing practices and achievement motivation appearing in folk tales. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (2nd ed.) New York: Holt, 1952. Pp. 243-249.
- MCCLELLAND, D. C., RINDLISBACHER, A., & DECHARMS, R. Religious and other sources of parental attitudes toward independence training. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Croft, 1955. Pp. 389-397.
- MCGUFFEY, W. H. *New fourth eclectic reader*. New York: Wilson, Hinkle, 1857.
- NIETZ, J. A. *Old textbooks*. Pittsburgh: Univer. Pittsburgh Press, 1961.
- PARKER, R. G., & WATSON, J. M. *The national second reader*. New York: Barnes, 1857.
- RIESMAN, D., GLAZER, N., & DENNEY, R. *The lonely crowd*. New Haven: Yale Univer. Press, 1950.
- ROSTOW, W. W. *The stages of economic growth*. Cambridge: Cambridge Univer. Press, 1960.
- SIEGAL, S. S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- STRAUSS, M. A., & HOUGHTON, L. J. Achievement, affiliation, and cooperation values as clues to trends in American rural society, 1924-1958. *Rural Sociol.*, 1960, 25, 394-403.
- UNITED STATES DEPARTMENT OF COMMERCE. *Historical statistics of the United States colonial times to 1957*. Washington: 1960.
- WEBER, M. *The Protestant ethic*. (Trans. by T. Parsons) New York: Scribner's, 1930.
- WHYTE, W. H. *The organization man*. Garden City: Doubleday, 1956.
- WINTERBOTTOM, M. R. The relation of need for achievement to learning experiences in independence and mastery. In J. W. Atkinson, (Ed), *Motives in fantasy, action, and society*. New York: Van Nostrand, 1958. Pp. 453-478.

(Received February 6, 1961)



## THE EFFECTS OF PROVIDING KNOWLEDGE OF RESULTS UPON THE PERCEPTUAL DEFENSE EFFECT<sup>1</sup>

MICHAEL J. GOLDSTEIN AND SAMUEL Z. HIMMELFARB

*University of California, Los Angeles*

Recent research on the role of motivational variables in perception has emphasized the need for a closer analysis of the factors operating in a perceptual recognition experiment. The traditional assumption that differences in perceptual accuracy for different classes of stimulus words or pictures reflect the exclusive contribution of the perceptual system has been seriously questioned by a number of writers (Eriksen, 1958; Goldiamond, 1958; Goldstein, 1962). The presence of response biases in articulating the indicator of perception, which exist independent of the presence of a discriminative stimulus, requires that the experimental arrangement for studying perception provide for the statistical or experimental isolation of these response biases from behavior truly under the control of a perceptual stimulus.

A second factor, which has received less attention than the concept of response bias, is the absence of feedback in the perceptual recognition situation. In most experiments, each time the subject makes a guess as to the meaning of an impoverished visual stimulus, he is not provided with any sort of veridical criterion against which to compare his guess. Thus, individual differences in criteria for reporting sensory inputs can be utilized to reinforce certain recognition guesses as being correct and to inhibit other guesses due to their being judged incorrect. This freedom of reinforcement can provide an opportunity for response biases to undergo various sorts of transformations depending upon the subject's reaction to not receiving

perceptual feedback. Once again we might misinterpret changes in perceptual accuracy over trials as being due to changes in discriminative ability rather than as systematic changes in response bias due to failure to receive feedback.

Interest in the present problem grew out of an earlier study (Goldstein, 1962) in which provisions were made to observe the perceptual defense effect over trials. Two things were noted in this study: the absolute magnitude of the perceptual defense effect *increased* over trials and the probability of using anxiety words as recognition guesses, which was lower initially than the probability of using neutral words as guesses, *decreased* over trials. The present report is concerned with exploring the degree to which these systematic changes in accuracy and response bias are a result of the lack of feedback from the experimenter during the process of measuring the subject's perceptions.

Specifically, it was predicted that the introduction of knowledge of results early in the series of perceptual responses would lead to: (a) a gradual equalization of the probability of using anxiety words and neutral words as recognition guesses and (b) that these changes in response bias will be paralleled by a similar equalization over trials of perceptual accuracy scores for anxiety words and neutral words.

### METHOD

*Subjects.* Subjects were 20 UCLA summer session students in a course in psychological aspects of aging given by the senior author. The subjects were older on the average than college sophomores typically used in psychological research, the majority falling between 20-25 years of age. The subjects were run by the junior author and every attempt was made to mask the senior author's interest or participation in the study. Five male and five female subjects were randomly assigned to each experimental group.

<sup>1</sup> The present study was supported by a grant from the UCLA Faculty Research Committee. The paper was worked on by the senior author while a Fulbright research fellow to the Psychological Laboratory, University of Copenhagen, 1960-61. An abridged version of this paper was presented at the Western Psychological Association meetings, April 1960 at San Jose, California.



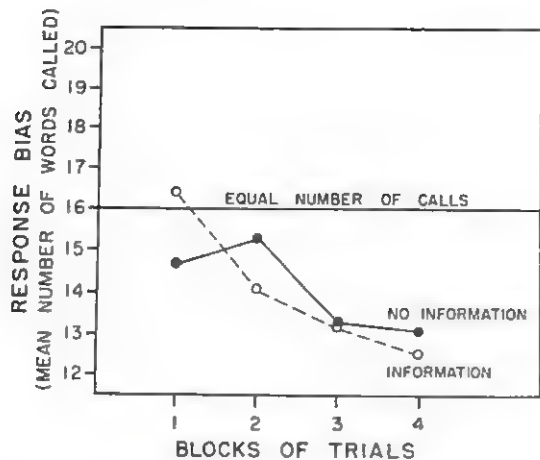


FIG. 1. Mean number of anxiety words used as recognition guesses broken down by information group and block of guessing trials.

**Selection of anxiety arousing and neutral words.** The method of selecting four anxiety words and four neutral words for each subject is identical to that reported in an earlier experiment (Goldstein, 1962). Following the selection session, the subject was scheduled to return within 48 hours.

**Perceptual session.** On the second experimental day, two things were done. First, a measure of the subject's visual sensitivity was obtained to determine an exposure time at which the subjects would be approximately 25% correct in his recognition guesses. Second the actual perceptual defense task was run. The method for determining visual sensitivity is also described elsewhere (Goldstein, 1962).

**Perceptual defense measurement.** Following the determination of visual sensitivity, the subject was told that the stimuli to be flashed in the Gerbrands tachistoscope were to be drawn from a set of eight words. The subject was shown the list of eight words, drawn from his word association test, and was given one minute to familiarize himself with them. Following this the subject was instructed:

This is a study dealing with perception. Here is a list of eight words which will be presented one at a time, in a random order, in this machine. You are to look into the eyepiece and judge which of the eight stimuli is present on each trial by referring to your list. The words will be presented fairly rapidly so that at times it will be fairly difficult for you to see the words and you will frequently have the feeling that you are guessing. Don't be disturbed by this, just guess on every trial.

Following this, the subject was shown the fixation point in the tachistoscope and was instructed to expect the stimuli about this point. After the completion of the instructions, stimulus words, typed on 3 × 5 cards, were exposed one at a time in the tachistoscope according to a prearranged random

order for 128 exposures. The only restriction on the random order was that each of the eight words had to appear equally often in each block of 32 trials, a provision which permitted a later analysis of the data over blocks of 32 exposures. All 128 exposure trials were run at the same setting, which was determined in the manner described above.

**Introduction of the knowledge of results factor.** Prior to the perceptual part of the experiment subjects were randomly divided into two groups, an information group and a no information group. For the first 32 stimulus exposures in the perceptual defense situation neither group of subjects received any information concerning the correctness of their responses. Following the first block of 32 trials, subjects in the Information group were told "Right" every time they guessed correctly. No transition to the information stage was made by the experimenter who merely started providing information on the thirty-third trial with no explanation to the subject. The No Information subjects received no feedback throughout the experiment.

## RESULTS

**Analysis of changes in response bias.** The first hypothesis tested was that the introduction of knowledge of results would result in a gradual equalization of the rate at which anxiety words and neutral words would be used as recognition guesses. To study this, each subject's record was scored for the number of times anxiety words were used as recognition guesses. To observe changes over trials, the total of 128 recognition trials was broken down into four blocks of 32 trials each. Equal

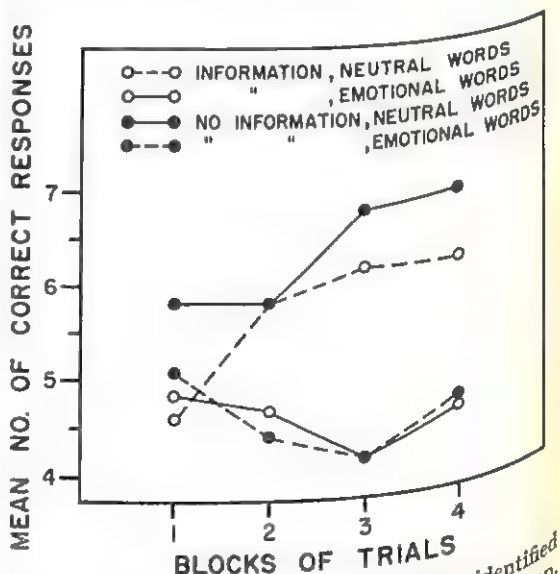


FIG. 2. Mean number of words correctly identified broken down into type of word, information group, and block of guessing trials.

usage of anxiety words and neutral words would be reflected in a score of 16 per block. Figure 1 presents the mean number of times anxiety words were used as recognition guesses in the Information and No Information groups. Inspection of Figure 1 indicates that the probability of using anxiety words as recognition guesses *decreases* in both groups at a very similar rate. The curve for the Information group clearly runs counter to our prediction. The response bias data were analyzed statistically using an analysis of variance of linear trend, in the manner suggested by Grant (1956). This analysis revealed an insignificant  $F$  ( $F < 1.00$  for  $df = 1/18$ ) for treatment groups which indicates that the linear component of the slope in both curves in Figure 1 do not differ. Knowledge of results has no noticeable influence upon response bias.

**Analysis of accuracy scores.** The second hypothesis was that accuracy scores for anxiety words and neutral words in the Information group would also become more similar over trials. These data are represented in Figure 2. Here the mean number of correct recognitions is presented, broken down by the type of stimulus word, treatment group, and block of trials. Although both treatment groups were treated identically in the first block of 32 trials, Figure 2 suggests an inequality in accuracy on the first block between the two groups. To determine if these first block scores differed significantly from each other, an analysis of variance was done on the first block scores alone. The  $F$  between groups (Information vs. No Information) was 1.59 which is not significant at the .05 level for  $df = 1/18$ . The  $F$  for the interaction between treatments and type of stimulus word was 2.02 which also fails to reach statistical significance (for  $df = 1/18$ ) indicating that the four first block scores did not differ significantly among each other. Since this was the case, the analysis of trends over trials used the scores from all four blocks of trials. Further inspection of Figure 2 also fails to suggest a decrease in the magnitude of the perceptual defense effect over trials. The slopes of the curves for the Information and No Information conditions, although starting from slightly different points of origin, are

TABLE 1  
ANALYSIS OF VARIANCE OF LINEAR SLOPE FOR  
INFORMATION AND NO INFORMATION  
GROUPS

Source	SS	df	F
Between groups:	909.50	20	
Mean	24.02	1	—
Information vs. No Information	65.02	1	—
Error	820.46	18	
Within groups:	1034.48	20	
Type of word	308.02	1	8.101*
Word $\times$ Treatment	42.02	1	
Error	684.44	18	

\*  $p < .05$ .

very similar. An analysis of variance of linear trend was also performed on the accuracy data and the results of the analysis are presented in Table 1. The only significant sources of variance in Table 1 is the type of word (anxiety word vs. neutral word), which means that the slopes of the curves for the neutral words differ significantly from the slopes of the curves for the anxiety words. A further look at Figure 2 reveals that it is the curves for the neutral words in both treatment groups which show a sharp inflection upward, while the anxiety word curves show only a slight downward trend. These data taken in conjunction with the previously reported response bias data suggest that the increase in accuracy for neutral words is accomplished by sharply increasing the rate at which these words are used as recognition guesses, a situation which appears to have little effect upon the accuracy level for anxiety words.

### DISCUSSION

The main hypothesis of the present study is clearly not supported; knowledge of results does not modify accuracy or response bias in a perceptual defense experiment. Evidence still exists from the progressive shifts in accuracy and response bias for a self-reinforcing process which modifies accuracy and probability of response, but the locus of the reinforcement is not clear. It is obviously not a simple case of the subject not having a veridical criterion against which to compare his recognition guesses.

Some clarification may come from attempting to unravel the direction of the causal re-



relationship between the changes in accuracy and the changes in response bias noted over trials. Two possibilities exist: The accuracy changes are primary and the changes in the response bias are secondary effects of the accuracy changes. The changes in response bias are primary and the changes in accuracy are an artifact of the changes in probability of usage of the two classes of words as recognition guesses. Evidence from an earlier study (Goldstein, Himmelfarb, & Feder, 1962) tend to support the second hypothesis. In this study, the perceptual recognition experiment was changed so that the subject indicated his perception by locating the stimulus in space, rather than by articulating the stimulus word, as was the case in the present study. The use of the spatial forced-choice indicator prevents response bias from affecting perceptual recognition scores. In the forced-choice study, no changes in accuracy over trials were noted for either class of stimulus word. This finding suggests that the changes in accuracy noted in the present study are an artifact of systematic shifts in response bias. The subject is restricting his recognition guesses progressively to that subset of responses termed neutral words, and is, in this fashion, increasing the probability of a chance congruence in accuracy for the neutral words.

If the hypothesis that it is response bias and not accuracy which shifts over trials, should be supported by other research, then how can we account for this type of shift? Two hypotheses are tenable, one which emphasizes the presence of a perceptual stimulus and a second which emphasizes response variables exclusively. The first hypothesis centers about the type of implicit verbal reinforcement which the subject is believed to administer after using each class of words as recognition guesses. It has been shown (Goldstein, 1962), that anxiety words have a lower probability of evocation than neutral words, when a blank stimulus is flashed in a tachistoscope. It is conceivable that the subject feels less subjective confidence after using a low probability word as a response than when he uses a high probability word. These differences in subjective confidence could affect the rate at which the

subject uses these two classes of words as responses over a series of trials. The second possibility is that this behavior is not under control of the stimulus at all. Let us forget for a moment that a real perceptual stimulus was presented to the subject and assume that the task assigned to him is to label some objects. He is given two classes of labels and told that he is to use them to label different objects. One class of labels has been associated with anxiety in the past while the other has not. What we then have is a classic avoidance learning paradigm. The subject has the choice of emitting an anxiety linked word or a neutral word on each trial. The anxiety linked word is painful for him to emit, so therefore he emits the neutral word whenever possible. Each time he uses a neutral word as a response, he is avoiding using an anxiety word, a situation closely paralleling many avoidance learning experiments. If we assume that not using an anxiety linked word is reinforcing of his already existing avoidance tendencies for that word, then each time the subject uses a neutral word, he is increasing the likelihood of using that word again because it is an anxiety reducing avoidance response. Thus, the progressive increase in the use of neutral words and decline in the use of anxiety words is explainable from an avoidance conditioning paradigm and does not even require the presence of a visual stimulus. Experiments, currently underway, using blank stimulus groups are directed at choosing between these two hypotheses. Should the same sort of progressive changes in response bias show up with no true discriminative stimulus present, then Hypothesis 2, which emphasizes reinforcement of intra-verbal habits, would be reasonably well supported.

Regardless of which hypothesis is supported, it should be apparent from the present research that the perceptual recognition experiment does not merely reveal stable reaction tendencies, but is itself a dynamic situation in which sequences of responses are modifying pre-existing reaction tendencies.

#### SUMMARY

The present paper reported a study designed to observe the perceptual defense



effect over trials. It was hypothesized that the introduction of knowledge of results would result in an equalization of the probability of using anxiety words as responses and in the accuracy for neutral and anxiety words. Both hypotheses failed to be confirmed. The probability of using anxiety words as recognition guesses *decreased* over trials and the accuracy curves diverged over trials.

Discussion centers about specifying the factors responsible for the systematic changes observed over trials.

## REFERENCES

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation*:

1958. Lincoln: Univer. Nebraska Press, 1958. Pp. 169-227.
- GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, 55, 373-411.
- GOLDSTEIN, M. J. A test of the response probability theory of perceptual defense. *J. exp. Psychol.*, 1962, 63, 23-28.
- GOLDSTEIN, M. J., HIMMELFARB, S., & FEDER, WALDA. A further study of the relationship between response bias and perceptual defense. *J. abnorm. soc. Psychol.*, 1962, 64, 55-62.
- GRANT, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141-154.

(Received February 13, 1961)

## CRITIQUE AND NOTES

### SEATING ARRANGEMENT AND LEADERSHIP EMERGENCE

LLOYD T. HOWELLS AND SELWYN W. BECKER

*Graduate School of Business, University of Chicago*

In task oriented discussion groups a leader will tend to emerge as a result of interaction among the group members. Bavelas (1950) and Goldberg (1955) have concluded that the size of the group and the type of communication pattern used are determiners of leadership emergence. Heise and Miller (1951) presented evidence to support the hypothesis that the most centrally located member of a group is likely to emerge as leader. Bales (1953) also found after analyzing communication patterns and leadership that "the communication pattern tends to 'centralize' . . . around a leader through whom most of the communication flows." Excluding Bales' analysis the experimental conditions in these studies were characterized by controlled communication patterns where the number of channels and the direction of communication were fixed.

With free communication possible, i.e., where all channels are available to all members of the group, Festinger, Schacter, and Back (1950) found that communication among individuals tends to be maximized between individuals positioned opposite to one another and conversely, that there is less communication between people placed side by side. A similar conclusion was drawn by Steinzor (1950) who stated that "the greater . . . the expressive stimulus value a member of a group has for others the more nearly opposite he sits from one in a circle" (p. 554). Because his subjects were seated in a circle, it was also possible for Steinzor to conclude that a participant who had a greater mean seating distance from all the other participants would be likely to attain higher leadership status.

Bass and Klubeck (1952) partially tested Steinzor's (1950) conclusions by arranging subjects on each side of a V shaped arrangement. With this arrangement the mean seating distances varied and following Steinzor it was expected that the subjects on the open end of the V would more frequently emerge as leaders. The data did not conform to this expectation and Bass concluded that the "particular seat a person occupied was of negligible importance in determining a participant's tendencies to attain leadership status . . ." (p. 727).

From the Festinger et al. (1955) and Steinzor (1950) results one could deduce that the communications of a small group sitting at a table will tend to flow "across" rather than "around" the table. Further, from the Bavelas' (1950) and Bales' (1953) results it can be stated that the individual who controls communication, i.e., the individual who sends and receives more messages, will emerge as leader.

If these two conjectures are correct then, despite Bass' (1953) failure to find the effect, seating arrangements can affect communication flow which, in turn, will influence leadership emergence. Accordingly, the general hypothesis to be tested in the present study is that *seating position, by influencing the flow of communication, is a determiner of leadership emergence.*

#### METHOD

*Subjects.* One hundred undergraduate men enrolled at the University of Chicago volunteered in order to earn \$1.00 as experimental subjects. Twenty groups of five were constituted by random selection from the pool of 100 subjects with the restriction that all members of a group be unable to recognize the names of the remaining group members.

For purposes of identification as each group member appeared for his evening session he was given a colored wrist band (Blue, Red, Green, White, or Yellow) to which he had been randomly assigned and instructed to sit in the chair with the matching color. The colors on the chairs were rotated so that over the 20 sessions each number was assigned to each chair position an equal number of times. The chairs were arranged so that two were evenly spaced along one of the long sides of a 6' x 3' table and three were evenly spaced on the opposite, long side of the table. The subjects were informed that the purpose of the experiment was to investigate the effects of various communication patterns on creativity after which the instructions pertinent to the task were given.

The instructions included an explanation of the task, anagrams, and the rules for evaluating the group's production. The subjects were told that each word was worth a number of points and that some words were worth more than others, and furthermore that as they generated words it would be possible to establish a hypothesis about the kinds of words which have greater value. They were

also told that a clock was set at random so that the alarm signaling the end of the session would ring sometime between 15 and 45 minutes after they started working. In reality all groups worked for 25 minutes. At the alarm the group was required to submit the product of their effort on one sheet of paper.

The task selected, and the conditions imposed on its completion, were designed to stimulate group interaction and to enhance the chances for development of group structure. That is, it was possible for one member to designate another as recorder so the requirement of submitting "one sheet representing the group's effort" would be satisfied. Decisions had to be made on the division of time and labor with regard to generating words and looking for the clue.

When the alarm rang ending the work session, questionnaires were distributed and completed by the group members. The questionnaires required each subject to rank the other four subjects, but not himself, on 15 personality characteristics and also to compare each subject with all the other subjects, again excluding himself, for leadership qualities demonstrated during the session. Since it has been shown that nominations by group members are more reliable leadership measures than are observer ratings or other leadership criteria (Carter, Haythorn, & Howell, 1950), the subjects' leadership ratings were taken as the criteria of leadership. Where there was a 4-1, or 3-2 vote favoring a single person, he was considered the leader of the group. Any other vote was deemed inconclusive and the leader was assumed to be on the side of the table mitigating against the hypothesis.

The procedures provided for two subjects to face three subjects across a table and engage in a problem solving task. Based on the theoretical considerations described above it was expected that communication would flow across rather than around the table. Thus, it was expected that each of two subjects would exert influence on three subjects and each of three subjects would exert influence on two subjects. Leadership ratings made by the members of the group provided data for a test of the following hypothesis:

*A greater number of leaders than would be expected by chance will emerge from the two-seat side of the table.*

## RESULTS AND DISCUSSION

The analysis of the data was based on the assumption that each person in the group had an equal chance of being voted leader. As such, it would be expected that 8 leaders would emerge from the two-seat side of the table and 12 from the three-seat side. Since the observed results are in the predicted direction and differ significantly from those expected by chance it is concluded that seating position influences com-

TABLE 1

NUMBER OF LEADERS EMERGING FROM TWO- AND THREE-SEAT SIDES OF TABLE

Two-seat side	Three-seat side
14	6

Note.— $\chi^2 = 7.5$ ,  $p < .01$ .

munication and so is a determinant of leadership emergence in a task oriented group.

Given these data it seems possible to reconcile the differences between Steinzor's (1950), Bass and Klubeck's (1952), and the present study. Steinzor found that a person sitting opposite another around a circle provided more stimulus for that other and also that the largest mean seating distance was a determiner of leadership. That the second conclusion may be viewed as an artifact of the circular arrangement of subjects seems possible, especially since Bass and Klubeck failed to support the conclusion with subjects arranged around a V shaped arrangement. If "oppositeness" rather than seating distance is important then Bass and Klubeck could have expected no differences in their situation, since no subject had an advantage in the number of other subjects he faced. Differences could be expected in the situation reported here, i.e., two subjects opposite three subjects, for each of two could possibly influence each of three.

## SUMMARY

Based on the proposition that seating position effects the flow of communication which, in turn, effects leadership emergence 20 groups of five subjects performed a problem solving task and then rated the group members on exhibited leadership. The seats were arranged so that two subjects were opposite three subjects. A greater number of leaders than would be expected by chance emerged from the two-seat side of the table, thus, lending support to the hypothesis.

## REFERENCES

- BALES, R. F. The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. A. Shils (Eds.). *Working papers in the theory of action*. Glencoe, Ill.: Free Press, 1953. Pp. 111-161.
- BASS, B. M., & KLUBECK, S. Effects of seating on leaderless groups. *J. abnorm. soc. Psychol.*, 1952, 47, 724-727.
- BAVELAS, A. Communication processes in task-oriented groups. *J. Acoust. Soc. Amer.*, 1950, 22, 725-730.



- CARTER, L. F., HAYTHORN, W., & HOWELL, M. A further investigation of criteria of leadership. *J. abnorm. soc. Psychol.*, 1950, **45**, 350-360.
- FESTINGER, L., SCHACHTER, S., & BACK, K. *Social processes in informal groups: A study of a housing project*. New York: Harper, 1950.
- GOLDBERG, S. C. Influence and leadership as a function of groups structure. *J. abnorm. soc. Psychol.*, 1955, **51**, 119-122.
- HEISE, G., & MILLER, G. Problem solving by small groups using various communication nets. *J. abnorm. soc. Psychol.*, 1951, **46**, 327-335.
- STEINZOR, B. The spatial factor in face to face discussion groups. *J. abnorm. soc. Psychol.*, 1950, **45**, 552-555.

(Received January 17, 1961)

## THE EFFECTS OF PATIENTS' RELATIONSHIPS WITH PEERS AND STAFF ON THEIR PSYCHIATRIC TREATMENT PROGRAMS<sup>1</sup>

E. ROBERT SINNETT

University of Kansas

AND DANA B. HANFORD

Northern State Hospital, Sedro-Woolley, Washington

Psychiatric treatment of the individual patient is influenced by social-psychological factors other than his need for therapy and the type and severity of his condition as denoted by psychiatric judgments of diagnosis and prognosis. Hollingshead and Redlich (1958), Hunt (1959), and others have provided convincing evidence that membership in a given social class and the characteristic psychological orientation associated with that class are substantially related to the treatment of the psychiatric patient.

The present study was an exploratory probe of the relationship between the affectional responses elicited by a patient from his ward physician and his peers and the type of treatment program he received. One objective was to investigate whether a physician's degree of like or dislike for a patient influenced the patient's assignment to individual psychotherapy.<sup>2</sup> Similarly, by means of sociometric data, the ties a patient had to his peers were examined for their possible relationship to his receiving individual psychotherapy. For purposes of comparison, however, both socioeconomic indices and psychiatric judgments of diagnosis and prognosis were included. Thus, three sets of variables were analyzed in relation to a patient's assignment to individual psychotherapy: a patient's relationships to his fellow patients and to his ward physician; socioeconomic indices (educational and occupational levels); and psychiatric judgments of diagnosis and prognosis.

### METHOD

**Subjects.** The subjects were 85 male veteran psychiatric patients on an open ward intensive treatment service. This service consisted of four 24-bed wards which were parallel, quite close to each other, and

<sup>1</sup> Research was done at Veterans Administration Hospital, Topeka, Kansas.

<sup>2</sup> Patients did not receive individual psychotherapy from their own ward physician. Hence this study is not concerned with the liking of the therapist for his patients but with the affective responses accorded the patient in his hospital environment by peers and ward physician.

joined by a common corridor. Social contacts among patients were not limited to members of a single ward since most activities both planned and informal were common for all four wards.

The diagnostic composition of the group was: psychoses, 48%; neuroses, 36%; personality disorders, 10%; and neurological disorders, 6%. The age range was from 19 to 67 with a mean of 37.5 years. The mean rating of educational level (Warner, Meeker, & Eells, 1949) was slightly lower than completion of high school. The mean rating of occupational level (Warner et al., 1949) was at the level of skilled workers.

The median length of current, continuous hospitalization for the subjects at the time of the study was 8 months, and the range was from 1 to 123 months. In order to assure each patient a reasonable time unit in which relationships might be established, a minimum of 1 month current hospitalization was used as a criterion for inclusion in the study. Subsequent analysis of the data showed that sociometric status was unrelated to length of hospitalization ( $\chi^2 = 3.97$ ,  $df = 3$ ,  $ns$ ).

**Sociometric questionnaire.** The sociometric questionnaire constructed for this study utilized free choice. That is, in responding to eight questions such as, "With whom would you like to be in group as therapy?" and "With whom do you prefer not to associate?" a subject was free to nominate as many or as few fellow patients as he chose. Choices were not restricted to a particular ward or service.

Ninety-three percent of the subjects returned their questionnaires. Twenty-one percent of the total group returned questionnaires signed but without responses. Failure to submit a questionnaire or returning a blank questionnaire was originally considered a basis for excluding individuals from the study. However, examination of the data showed that both these groups received statistically fewer choices from their peers ( $\chi^2 = 17.01$ ,  $df = 2$ ,  $p < .001$ ). Therefore, it was decided to regard both nonresponding classes as expressing meaningful social behavior in their nonresponse rather than to bias the sample by excluding these less popular subjects. The number of rejections received from peers did not differentiate these classes of nonrespondents from those completing the questionnaire ( $\chi^2 = .62$ ,  $df = 2$ ,  $ns$ ).

**Interview and rating procedures.** The patients' ward physicians, who were second or third year

TABLE 1

RELATIONSHIP BETWEEN SOCIOMETRIC CATEGORY AND ASSIGNMENT TO INDIVIDUAL PSYCHOTHERAPY

Sociometric category	Treatment		N
	No individual therapy	Individual therapy	
Popularity			
High	9	16	25
Average	23	9	32
Rejects	6	2	8
Isolates	14	6	20
Totals	52	33	85

Note.— $\chi^2 = 9.50$ ,  $df = 3$ ,  $p < .05$ .

psychiatric residents, were interviewed for information about the patients' treatment programs and their progress in treatment. The four psychologists who did the structured interviews with the physicians felt that a condition of good rapport existed since most of them had regular, sustained contact with the residents as part of their clinical duties. At the time of the interviews the residents had known their patients for from 1 to 6 months. These 13 physicians were also asked to rate their degree of liking or disliking for each of their patients on a seven-point scale.

Diagnosis from the final summary and prognosis from the initial report of examination (based on a 2-week evaluation) as well as information about length of hospitalization and milieu treatment program were obtained from the patients' records.

## RESULTS

The statistical analyses used chi square throughout. Three approaches were used to analyze the sociometric data. First, individuals were categorized as relatively isolated; relatively rejected; popular, or of medium popularity on the basis of the pattern of choices made and received; rejections made and received; and the number of mutual or reciprocal choices (i.e., persons who chose one another). Distributions of each of these five variables were dichotomized or trichotomized (where feasible) in order to classify subjects. For example, the isolates, as defined in this study, were those who made and received relatively few choices or rejections and had few mutual choices. The rejected individuals were those who received a relatively high number of rejections. It must be kept in mind that these categorizations are not absolute. For example, it happened that no patient who received a high number of rejections received *no* choices. That is, the most rejected patients were not unpopular but rather, controversial. Therefore the data were analyzed not only by sociometric categories, but also, as a second method of analysis,

by the component scores (number of choices made, rejections received, etc.) from which the categories were derived.

The set of sociometric categories was found to be significantly related to whether or not a patient was given or had been given individual psychotherapy (see Table 1). Relatively popular patients were more likely to receive individual psychotherapy than less popular patients.

Analysis of the sociometric questionnaire by component scores revealed that patients who were treated by individual psychotherapy received more choices from their peers ( $\chi^2 = 3.84$ ,  $df = 1$ ,  $p < .05$ ); they also made more choices ( $\chi^2 = 5.12$ ,  $df = 1$ ,  $p < .05$ ) and made more rejections ( $\chi^2 = 7.61$ ,  $df = 1$ ,  $p < .01$ ). They also received more rejections from their peers but not to a significant degree ( $\chi^2 = 2.39$ ,  $df = 1$ ,  $p < .20$ ). As an outcome of these analyses a factor of affect intensity of peer relationships appeared to be related to receiving individual psychotherapy. In order to construct a combined measure of intensity of ties to peers as a third approach to data analysis, number of choices made, number of choices received, number of rejections made, and number of rejections received were dichotomized at their medians. Scores for each of these variables were treated as follows: values above the median were given a weight of 1; those below the median, 0; and these weights were summated for each subject over the four variables. The resultant affect intensity score was found to be positively related to receiving individual psychotherapy ( $\chi^2 = 5.27$ ,  $df = 1$ ,  $p < .05$ ). Thus, both expressing and eliciting positive or negative feelings to or from one's peers were associated with receiving individual psychotherapy.

Physician liking was positively related to receiving individual psychotherapy; i.e., patients liked by their ward physician were more likely to receive this form of treatment (see Table 2).

TABLE 2

RELATIONSHIP BETWEEN BEING LIKED BY ONE'S PHYSICIAN AND ASSIGNMENT TO INDIVIDUAL PSYCHOTHERAPY

Physician liking	Treatment		N
	No individual therapy	Individual therapy	
Liked by physician	34	31	65
Not liked by physician	18	2	20
Totals	52	33	85

Note.— $\chi^2 = 9.13$ ,  $df = 1$ ,  $p < .01$ .



TABLE 3

RELATIONSHIP BETWEEN SOCIOMETRIC CATEGORY  
AND DIAGNOSIS

Sociometric category	Diagnosis <sup>a</sup>				N
	Brain damaged	Psychotics	Personality disorders	Neurotics	
Popularity					
High	1	10	4	10	25
Average	1	16	4	11	32
Rejects	2	5	0	1	8
Isolates	1	9	1	9	20
Totals	5	40	9	31	85

Note.— $\chi^2 = 3.42$ ,  $df = 3$ , *ns*.  
<sup>a</sup> Because of many small cell frequencies, the categories were combined as follows for the  $\chi^2$  test: brain damaged subjects were excluded and neurotics and personality disorders were combined.

It should be noted that the practice in the hospital setting where the study was conducted was for ward physicians *not* to take their own patients in psychotherapy. Assignment to individual psychotherapy was characteristically a product of a joint decision of a ward physician and his supervisor, or an outcome of a staff conference where the resident presented his psychiatric case study of a particular patient.

Although the relationship between physician liking and receiving psychotherapy is open to the interpretation that therapy made patients more likeable rather than being a factor in selection for treatment, there is some evidence that would tend to be incongruent with the former interpretation—namely, the fact that among the 33 subjects who received psychotherapy, physician liking showed a low, *negative*, insignificant relationship ( $r = -.16$ ) with number of months in psychotherapy.

An additional finding from this study supports the interpretation that patients who are ignored by or isolated from their peers by their own choice receive a less intensive treatment program. Physician's prescriptions for assignments of patients to activities tended to include fewer stated treatment objectives for isolates and also for rejects than for patients of higher popularity (the Kruskal-Wallis  $H$  test yielded a  $\chi^2$  value of 6.38,  $df = 3$ ,  $p < .10$ ).

Consistent with findings of a previous study using subjects drawn from the same population (Sinnott & Hammersley, 1957), prognosis on admission was unrelated to receiving individual psychotherapy ( $\chi^2 = 0.11$ ,  $df = 1$ , *ns*). In the earlier study diagnosis was also unrelated to receiving individual psychotherapy. However, in contrast with the earlier finding, the present study showed that diagnosis and receiving individual psychotherapy were related: neurotics were more likely to receive this form of treatment ( $\chi^2 = 10.27$ ,  $df = 3$ ,  $p < .02$ ).

Although the population receiving individual psychotherapy had a higher educational level ( $\chi^2 = 5.56$ ,  $df = 1$ ,  $p < .02$ ), occupational level was not so related in this sample ( $\chi^2 = 0.32$ ,  $df = 1$ , *ns*). The findings from the previous sample (Sinnott & Hammersley, 1957) were that both the more highly educated and those who had held higher level jobs more often received individual psychotherapy.

Although one might anticipate that degree of isolation, rejection, and popularity would be related to diagnosis and prognosis—for example, that schizophrenics might predominate among the isolates—no such relationship was found (see Table 3). Similarly, diagnosis and prognosis were not related to being liked by one's physician

( $\chi^2 = 1.05$  and  $0.00$ ,  $df = 3$  and  $1$ , respectively) even though in some cases the three judgments of liking, diagnosis, and prognosis might have been made by the same physician. Thus, it appears that no generalized halo effect confounded these judgments.

Somewhat in contrast to what has been found by Hollingshead and Redlich (1958), indices of socioeconomic level did not show a substantial relationship to physician liking. Being liked by one's physician was not related to occupational level ( $\chi^2 = 1.27$ ,  $df = 1$ , *ns*). However, there was a trend toward a positive relationship between being liked by one's physician and educational level ( $\chi^2 = 3.19$ ,  $df = 1$ ,  $p < .10$ ). The more highly educated patients were also more popular with their peers ( $\chi^2 = 8.59$ ,  $df = 3$ ,  $p < .05$ ). There was a trend toward a relationship between sociometric category and physician's regard for a patient: patients who are relatively popular with their peers tend to be liked by their physicians ( $\chi^2 = 6.29$ ,  $df = 3$ ,  $p < .10$ ). The relationship between physician liking and status among patient peers was explored further using the affect intensity measure described above. Expressing and eliciting positive and/or negative feelings to or from one's peers was positively related to being liked by one's physician ( $\chi^2 = 3.85$ ,  $df = 1$ ,  $p < .05$ ). Conversely, those who made fewer choices and rejections and were infrequently chosen or rejected by peers were less liked by their physicians, just as they were less often selected for psychotherapy.

The sociometric categories were independent of age ( $\chi^2 = 2.07$ ,  $df = 3$ , *ns*) and length of hospitalization ( $\chi^2 = 3.97$ ,  $df = 3$ , *ns*). Physician liking also was not related to length of hospitalization ( $\chi^2 = 0.03$ ,  $df = 3$ , *ns*).

## DISCUSSION

It appears that a relatively intense affective responsiveness leads to a patient being considered a good candidate for psychotherapy. The degree to which affective responsiveness is an attribute necessary to the conduct of psychotherapy and the degree to which there are simply personal preferences among psychiatrists for working with more responsive patients are matters that are open to question.

To date the relationships of sociometric measures and physician liking to receiving individual psychotherapy have not been cross-validated. However, the findings of Hollingshead and Redlich (1958, pp. 344-345), imply a positive association between liking and selection for psychotherapy. They show that physicians like better those patients from higher social classes and that patients from higher social classes are more likely to receive psychotherapy.

Reciprocal or mutual choice behavior among the subjects of this study was strikingly similar to the results obtained by Murray and Cohen (1959) in their open ward psychiatric patient group, even though somewhat different sociometric measures were utilized in the two experiments. They found 50% of their patients had one or more reciprocal positive choices and the corresponding figure for the present study was 39%. One or more mutual negative choices or rejections were found by Murray and Cohen in 6% of their cases as compared with 0% in this study.

Although previous studies (Davids & Parenti, 1958; McMillan & Silverberg, 1955; Murray & Cohen, 1959; Northway & Widgor, 1947) and a review (Lindzey & Borgatta, 1954) have shown less consistency than might be desired, generally subjects receiving relatively few choices or relatively many rejections have been judged on the basis of other measures to be more maladjusted. In this study, however, there was no association found between diagnosis or prognosis and sociometric measures.

Ordinarily in the appraisal of patients a physician's feelings toward the patient and the patient's relationships with his peers are subordinated to "objective" considerations of diagnosis and prognosis; nevertheless, the matrix of interpersonal relations appears to make itself felt in the patient's treatment program.

The psychiatric patient comes to the hospital for treatment rather than be treated in his home environment mainly in order that his social climate may be controlled so that he can achieve a perspective view of problems with himself and in his relationships with others. Yet, should he

manage to duplicate among his patient peers and professional staff the climate he has left at home, the effectiveness of efforts to modify his pattern of relationships with others may be less than is supposed.

The findings appear to lend support to the current argument that the whole nexus of patient relationships—the social environment of the mental hospital—ought to be examined and treated as well as the problems of individual patients.

## SUMMARY

The major finding of this research is that a patient's relationships with his peers and his physician relate to the type of treatment program he receives as well or better than indices of his social class or his psychiatric diagnosis or prognosis. Among a group of 85 male veteran patients, the subjects most often selected for psychotherapy were: patients who were most popular among their peers; patients whose affective ties (both positive and negative) with peers were strongest; and patients most liked by their ward physicians. Conversely, the patient who was neither "in with" nor a controversial figure among his fellow patients, who was either only average in popularity or isolated from his peers, was not often selected for psychotherapy.

## REFERENCES

- DAVIDS, A., & PARENTI, ANITA N. Personality, social choice, and adults' perceptions of these factors in groups of disturbed and normal children. *Sociometry*, 1958, 21, 212-224.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- HUNT, R. G. Socio-cultural factors in mental disorder. *Behav. Sci.*, 1959, 4, 96-106.
- LINDZEY, G., & BORGATTA, E. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. *Theory and method*. Cambridge: Addison-Wesley, 1954. Pp. 405-448.
- McMILLAN, J. J., & SILVERBERG, J. Sociometric choice patterns in hospitalized ward groups with varying degrees of interpersonal disturbances. *J. abnorm. soc. Psychol.*, 1955, 50, 168-172.
- MURRAY, E. J., & COHEN, M. Mental illness, milieu therapy, and social organization in ward groups. *J. abnorm. soc. Psychol.*, 1959, 58, 48-54.
- NORTHWAY, MARY L., & WIDGOR, BLOSSOM T. Rorschach patterns related to the sociometric status of school children. *Sociometry*, 1947, 10, 186-199.
- SINNETT, E. R., & HAMMERSLEY, D. W. The relationship between attributes of patients selected for psychotherapy and improvement measures. *Amer. Psychologist*, 1957, 12, 432. (Abstract)
- WARNER, W. L., MEEKER, MARCHIA, & EELLS, K. *Social class in America*. Chicago: Science Research, 1949.

(Received January 12, 1961)



## BIRTH ORDER, NEED ACHIEVEMENT, AND CONFORMITY

EDWARD E. SAMPSON

*University of California, Berkeley*

### *Experimental Situations*

Although Alfred Adler (1945) suggested the importance of an individual's ordinal position in the family, until recently, very little research had been conducted using ordinal position as an independent variable. A few years ago, Schachter (1959) published a book in which he reported the interesting results of a series of studies on ordinal position, motivation, and behavior in social or experimental situations. In his work, Schachter suggested the hypothesis that first born individuals were more anxious and affiliatively dependent than later born persons. This suggests that the first born individual should be more prone to influence by others; and, in fact, Schachter reports the findings of a study by Ehrlich that supports this hypothesis. Ehrlich found that first born males were less resistant in a social influence situation than later born males.

The data reported in this paper derive from a series of three experiments, two of them directly concerned with social influence, the third more directly concerned with the need for achievement (*n Ach*).<sup>1</sup> As a routine part of each of these experiments, birth order data were obtained from each subject and analyzed in order to test the relationships between birth order and conformity in two experiments and birth order and *n Ach* in one experiment.

### METHOD

#### *Birth Order*

Information was obtained about the individual's order of birth by a simple question on which each subject indicated whether he was the first born, second born, etc. The data were classified into two major groupings: first born versus later born. The former category consisted of first and only and first with sibs. The latter category consisted of persons second born or later. For the analyses of this study, such factors as age distance between self and sibs or sex of nearest sibs were not controlled because of the small *N*.

<sup>1</sup> The author wishes to thank John R. P. French, Jr. for his invaluable contributions in the first experiment, and Ernest Harburg for his collaboration in the third experiment, which he is now preparing for eventual publication.

There were two experimental situations which yielded data about conformity and birth order. The first situation, A, consisted of an audience-type influence experiment in which a hired assistant, playing the role of a debator, sought to influence the subject to adopt the position which he advocated on an attitude issue about Russia and the United States. The experiment used a  $2 \times 3$  design, with a condition of negative attraction towards the debator and one of neutral attraction. Within each of these, the debator, was a false-expert—that is, he based his influence on an assumed expertness which he actually did not possess; sought to control the subjects' behavior—that is, he told the subjects that he possessed the ability to completely determine their beliefs in this particular situation; or was neither—that is, the control groups. The measures of conformity consisted of the direction and distance of the subjects' change from their preinfluence to their postinfluence position.

The second experimental situation, B, consisted of a four-man group working in separate cubicles on a simple performance task. The task involved drawing circles with dots in the middle for a series of 10-second trials. During the course of the experiment, an individual assumed to be the leader of this four-man group gave instructions to the subjects to increase their performance. The leader promised rewards for conformity to his instructions to half the groups and threatened fines for non-conformity to the other half. The rewards and fines consisted of points given or taken away, the points being worth money. The group leader also gave each subject his evaluation of their performance on the task. Upon receiving his group leader's evaluation, each subject was asked to record his own personal evaluation on the same scale employed by the group leader. The measure of conformity in this situation consisted of the discrepancy between the subject's own personal evaluation and the group leader's evaluation of his performance. Thus, as with the conformity measure in Situation A, this involved a judgmental or cognitive conformity rather than an action conformity.

The third experimental situation, C, gave information about birth order and *n Ach*. It consisted of a digit-symbol task given to groups of persons premeasured on *n Ach* and test anxiety. The major concern in this situation was to determine differences in digit-symbol performance between persons motivated by the need to achieve as compared with persons motivated to avoid failure.



TABLE 1  
THE RELATION BETWEEN BIRTH ORDER  
AND THE NEED FOR ACHIEVEMENT

Total	Birth order		Total	Birth order	
	First	Later		First	Later
High n Ach	23	7	High Ach-Low TAQ	13	2
Low n Ach	15	16	Low Ach-High TAQ	6	9
$\chi^2 = 4.05^a$ $p < .05$			$p = .05$ by Fisher's exact test		
Males					
High n Ach	12	3	High Ach-Low TAQ	6	0
Low n Ach	9	7	Low Ach-High TAQ	3	3
$\chi^2 = 1.03^a$					
Females					
High n Ach	11	4	High Ach-Low TAQ	7	2
Low n Ach	6	9	Low Ach-High TAQ	3	6
$\chi^2 = 2.17^a$					

<sup>a</sup> Corrected for continuity.

Subjects

In Experimental Situation A, involving an attitude issue about Russia and the United States, there were 88 female subjects obtained from introductory psychology courses at the University of Michigan. In Experimental Situation B, involving four-man groups, there was a total of 116 male subjects obtained from recruits at the Coast Guard base in Oakland, California. The final set of subjects in Experimental Situation C consisted of 31 males and 30 females who were members of an undergraduate social psychology class at the University of Michigan.

Measurement of n Ach

n Ach was measured by the projective French (1958) Test of Insight.<sup>2</sup> Test anxiety was measured by the Mandler-Sarason test anxiety items (Mandler & Cowen, 1958). These measures were obtained only on persons in Experimental Situation C. The inclusion of the measure in test anxiety enables a further refinement to be made of n Ach and fear of failure. Subjects above the median on n Ach and below the median on test anxiety constitute one "pure" group in contrast to the subjects below the median on n Ach and above it on test anxiety. This is the type of breakdown suggested by Atkinson (1958; Atkinson & Litwin, 1960).

RESULTS <sup>3</sup>

Birth Order and n Ach

The analyses presented in Table 1 permit us to look at the achievement measure alone and the split involving both achievement and anxiety (Experimental Situation C). For the total sample

<sup>2</sup> Ernest Harburg coded the projective test data measuring n Ach.

<sup>3</sup> All statistical tests are two-tailed, as we have no firm basis for making specific directional predictions.

of both men and women, there is a significant tendency for first born persons to have a higher need for achievement than later born persons. This relation is supported in the finer breakdown, employing both the measure of achievement and the measure of anxiety. Both of these relationships are significant at the .05 level. The trend of the relationship is the same for both males and females taken as separate groups, but is not statistically significant ( $\chi^2 = 1.03$  for the males;  $\chi^2 = 2.17$  for the females). There is, however, a slight, nonsignificant indication that this relationship is stronger for females than for males.

Birth Order and Conformity

Experimental Situations A and B each involved an influence situation in which we could measure the conformity or resistance of first and later born persons. Table 2 presents the relevant data from Experimental Situation A. The statistical test is significant at the .10 level, with the direction of the relationship suggesting that first born persons have higher resistance to influence than later born persons.<sup>4</sup> This finding is directly opposite to that Ehrlich has reported by Schachter (1959), but his subjects were males.

Although the sample on which these conformity data were obtained is not the same as that on which the n Ach data were obtained, the groups being compared in each are female undergraduate college students at the University of Michigan.

<sup>4</sup> Because Experimental Situation A employed an experimental division of groups into those having a neutral debator and those having a disliked debator, we conducted an analysis of birth order and conformity within these two major groupings. For both those having a neutral debator and those having a disliked debator, there is an indication (not statistically significant, however) that later born persons are less resistant than first born persons. This finding is consistent with the findings reported in Table 2 for the total sample.

TABLE 2  
THE RELATION BETWEEN BIRTH ORDER OF FEMALES  
AND RESISTANCE TO INFLUENCE  
ON AN ATTITUDE ISSUE

	Birth order	
	First	Later
Low resistance to influence	18	29
High resistance to influence	24	17
$\chi^2 = 3.60, p < .10$		

TABLE 3

THE RELATION BETWEEN BIRTH ORDER OF MALES AND  
RESISTANCE TO ACCEPTANCE OF THE  
GROUP LEADER'S EVALUATION<sup>a</sup>

	Birth order		t
	First	Later	
Reward conditions	3.62	5.67	1.86*
Fine conditions	5.07	5.88	.88

<sup>a</sup> The measure reported is the mean discrepancy between the group leader's evaluation of the subject's performance and the subject's own evaluation of his performance.

\* Significant at .06 level.

and are, thus, a relatively homogeneous sample. If we look at these results on conformity and the results on *n* Ach, we note that they are consistent, but only the former approaches significance. That is, in general, first born females are more resistant in a social influence situation and tend to be higher in the need for achievement than later born females.

Table 3 presents the data from Experimental Situation B on birth order and conformity for the Coast Guard recruits under conditions of reward and fine. The direction of the differences in conformity is the same within reward and fine, with first born being *more* conforming than later born. The differences were significant, however, only in the reward conditions ( $p = .06$ ). A comparison of first and later born persons on their conformity in the six experimental conditions of the study supports the overall finding. In six out of six groups, the first born are more conforming than the later born. This finding is directly opposite to our preceding findings from Experimental Situation A, but is quite consistent with the Ehrlich findings.

### DISCUSSION

We have presented a series of results which suggest that (a) first born females are more resistant to social influence than later born females and (b) first born males are less resistant to social influence than later born males. There is also a slight, but nonsignificant, indication that first born females have a higher need for achievement than later born females and that this relationship between birth order and need for achievement is stronger for the females than for the males.

Although the influence situations on which these data were obtained were different, we shall make the working assumption that there is sufficient comparability to permit us to draw certain

conclusions about the important interactions between birth order, sex, and conformity.

In attempting to explain the original data relating birth order and conformity, Schachter (1959) made the assumption that the first born person who gets dethroned by later born sibs will be anxious about losing his position and losing the love of his parents, and would, thus, be oriented towards seeking attention, approval, and support from others. One could further assume that the first-and-only child—who by definition does not get dethroned by later sibs—is still involved in approval seeking behavior that has its source in the typically intense dependent relationship that parents form with an only child. The parents here may become the major source of support and approval, and anxiety over not pleasing them may become particularly intensified in the first-and-only child. Assuming this approval seeking, dependent behavior generalizes, one would expect the first born to conform more in a social influence situation than the later born. These assumptions are consistent with the Ehrlich data and with our data on males from Experimental Situation B, in which we find first born males less resistant (more conforming) to social influence than later born males.

However, we have reported another set of data—that involving the females from Experimental Situation A, which suggest that the first born female is more resistant to influence than the later born female. Thus, although the data from Experimental Situations A and B and our data and Ehrlich's are not strictly comparable, one could infer from the available data that the first born female is more independent and resistant to influence than the first born male. Accepting for the moment the validity of this inference, let us look at some possible assumptions one could make to explain it, as well as other research findings relevant to these assumptions.

One assumption relates early training in independence to the development of the need for achievement, and from this, to resistance to influence. Winterbottom (1958) reports a relationship between early training in independence and the development of need for achievement. Krebs (1958) offers further data on this point, although his correlation between independence training and *n* Ach was not significant. The picture becomes more complex when we look at findings relating the need for achievement with conformity. McClelland's (1953) analysis of the Asch data suggests a relation between high *n* Ach and independence from influence. Winterbottom reports that subjects high in *n* Ach less frequently



asked for help in a puzzle solving situation. Krebs finds that subjects with the most intense orientation towards achievement are the most resistant to opinion change. He also finds that there is greatest conformity when achievement is low and independence training is late. Similarly, Walker and Heyns (1956 unpublished) find an inverse relation between achievement and conformity. This rather neat picture is somewhat complicated by the work of Samelson (1958) who finds that *n* Ach and social approach interacted with the experimental treatment, producing *higher conformity* under some conditions for persons high in the need for achievement. In support of this finding, Burdick (1955) suggests that the independent, achievement oriented person should conform in situations in which he perceives conformity as leading to achievement.

It seems fair to conclude that in general there is a positive relationship between early training in independence and the strength of need for achievement, but the apparent positive relationship between high *n* Ach and resistance to influence may be altered under certain conditions, as when conformity leads to achievement.

Applying these assumed relationships between early training in independence, development of the need for achievement, and resistance to influence to the interpretation of our data, we suggest that the first born female is involved in rearing later born sibs (Koch, 1955), and that this involvement gives the first born female more training in independence than the first born male. This independence training for the first born female could lead to a higher need for achievement and eventually to greater resistance in a social influence situation.

Another related assumption one could make involves the differential significance and timing of independence training for males and for females. The nature of the family situation is such that the young girl is being introduced to her adult role at an earlier age than the young boy. "Helping mother around the house" is more a part of the girl's later role than it is a part of the boy's later adult role. Here we are assuming that the first born male, although expected to assume some forms of responsibility in helping around the house, generally is not expected to be as responsible a person in these home activities as is the female, because these do not form an important part of his later adult role. His later responsibilities will lie within an occupational role in which he is still too young to be expected to form responsible behavior patterns. Some of these assumptions are supported

by Koch's (1955) data in which she reports that generally girls are seen as more responsible than boys, and first born girls are rated higher in leadership (perhaps an indication of greater independence) than are first born boys. Koch, however, employed controls on sex of sibs and sib age separation, two conditions that were shown to interact significantly with birth order in affecting other variables such as leadership, ambition, etc.

We may now add to this picture another link in our chain of assumptions. Although the first born male and the first born female may be concerned over dependency and may be involved in seeking parental approval and love, the first born female initially has greater pressures than the male to develop responsibility as part of her role. She can handle both the needs to get parental approval and these pressures by the same activity: that is, by developing responsibility, she is receiving the approval of her parents. This does *not* mean that the male is not approved for exhibiting independent, responsible behavior. It does mean, however, that the parents expect greater responsibility at an earlier age from the female, apply greater pressures to the female to show it, and express greater approval when she exhibits such responsible, independent behavior. Consistent with this is Tuddenham's (1952) finding that in general girls are more approved by elders and peers than are boys.

What we are suggesting by our assumptions is that both the *significance* of the independence training and the *timing* of such training differ for the first born females as compared with first born males. Training in independence is more significant for the female and occurs at an earlier age than for the male.

Although one generally thinks of females as being more dependent than males, we are suggesting that the *first born* female is more independent and has a higher need for achievement than the *first born* male; and this independence results in greater resistance in a social influence situation on the part of the first born female as compared with the first born male. Without an analysis of birth order, the results from the work of Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner (1954) which suggests that females are more field dependent than males, and the work of Crutchfield (1955) which suggests that females conform more than males, cannot be directly integrated into our own scheme. It is interesting to note, however that Crutchfield reports that with a sample of older females, he gets a whole range of conformity and resistance, with the general trend showing the



females to be more resistant to influence than the adult male sample. It is also interesting to note that in a replication of Witkin's work (Gruen, 1955), there is no significant difference between males and females on some of the tests on which Witkin et al. originally reported a difference. Thus, the picture that females are more dependent and conforming than males is itself more complex than originally thought, as some studies, including our own, suggest that some females are more resistant than some males. It would be of interest to reanalyze the data of Crutchfield, Witkin, and Gruen, controlling the apparently important variable of birth order.

The above set of assumptions provides a source for deriving additional hypotheses which are testable in a more systematic research project. For example, one would expect (a) the first born female to have greater responsibility demands from her parents than the first born male, (b) parents to perceive early responsibility as being more important for the first born female child than for the first born male child, and (c) the first born female to be more resistant to social influence than the first born male. We have inferred support for this third hypothesis from our data. However, a strict test of that hypothesis would involve placing males and females into the same influence situation and measuring their resistance to influence.

#### SUMMARY

The combined results from three separate studies using three separate samples of subjects suggest the following three conclusions: first born persons have a higher need for achievement than later born persons; first born females exhibit greater resistance to influence than later born females; and first born males exhibit less resistance to influence than later born males.

These findings are taken to be consistent with a set of assumptions that the first born female is more significantly involved in independence training than the first born male. This early independence training produces a greater need for achievement and leads to greater resistance to influence for the first born females. First ordinal position for the male, on the other hand, produces greater affiliative dependency and leads to greater conformity in an influence situation.

#### REFERENCES

- ADLER, A. *Social interest: A challenge to mankind*. (Trans. by J. Linton & R. Vaughan) London: Faber & Faber, 1945.
- ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.
- ATKINSON, J. W., & LITWIN, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *J. abnorm. soc. Psychol.*, 1960, 60, 52-63.
- BURDICK, H. A. The relationship of attraction, need achievement, and certainty to conformity under conditions of simulated group atmosphere. Unpublished doctoral dissertation, University of Michigan, 1955.
- CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, 10, 191-198.
- FRENCH, E. G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 242-248.
- GRUEN, A. The relation of dancing experience and personality to perception. *Psychol. Monogr.*, 1955, 69(14, Whole No. 399).
- KOCH, H. L. Some personality correlates of sex, sibling position, and sex of sibling among five- and six-year-old children. *Genet. psychol. Monogr.*, 1955, 52, 3-50.
- KREBS, A. M. Two determinants of conformity: Age of independence training and achievement. *J. abnorm. soc. Psychol.*, 1958, 56, 130-131.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- MANDLER, G., & COWEN, J. E. Test anxiety questionnaires. *J. consult. Psychol.*, 1958, 22, 228-229.
- SAMELSON, F. The relation of achievement and affiliation motives to conforming behavior in two conditions of conflict with a majority. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 421-433.
- SCHACHTER, S. *The psychology of affiliation*. Stanford, Calif.: Stanford Univer. Press, 1959.
- TUDDENHAM, R. D. Studies in reputation: I. Sex and grade differences in school children's evaluation of their peers. *Psychol. Monogr.*, 1952, 65(1, Whole No. 333).
- WINTERBOTTOM, M. R. The relation of need for achievement to learning experiences in independence and mastery. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 453-478.
- WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

(Received January 23, 1961)

## THE RECALL OF DREAMS:

### ITS RELATION TO REPRESSION AND COGNITIVE CONTROL

FRANK M. LACHMANN, BENJAMIN LAPKIN, AND NORMAN S. HANDELMAN<sup>1</sup>

*Veterans Administration Regional Office, New York*

With the publication *The Interpretation of Dreams*, Freud (1959) opened the way for a more scientific approach to the study of dream phenomena. The psychoanalytic investigations which followed, and the inferences drawn, were based primarily upon individual case histories, difficult to replicate. Experimental studies have been for the most part nontheoretical in nature; Ramsey (1953) summarizes this voluminous empirical literature. Until recently the psychoanalytic and the experimental approaches were divergent paths, overlapping only accidentally. The experimental findings of Kleitman (1960), demonstrating the universality of the dream process, have provided a baseline for laboratory investigation along theoretical psychoanalytic lines. Dement and Kleitman (1957) and Goodenough, Shapiro, Holden, and Steinschreiber (1959) noted the occurrence of rapid conjugate eye movements and electroencephalographic activity at specific intervals during sleep which were found to coincide with dream activity. Even subjects who said they had never remembered dreaming prior to the study reported dreams when awakened during eye movement periods.

Dreaming seems clearly a universal phenomenon, an "intrinsic part of normal sleep" (Dement & Kleitman, 1957, p. 345). It further appears that there is a great uniformity in the pattern of dream activity among people. One would expect a comparable uniformity in the ability to remember dreams and yet, as the literature clearly shows, this is not the case. There are individuals who scarcely recall dreaming while others report dream activity almost every night. This has led to investigations to account for the differences between such divergent types of behavior. Schonbar (1959) found that subjects who recalled dreams more often than 25% of the

time (the high dream recall group) were more "intelligent" and experienced more "overt anxiety" than the nonrecall group (those who reported dreams once a month or less). In addition, she reported a high correlation between remembering dreams and the contentless recall of dreaming. From this she hypothesized that low dream recall is related to repression, which conforms to the general psychoanalytic theory (Freud, 1959) that forgetting dreams or failing to recall dreams is a function of repression.

Personality differences associated with the variations in the amount of dream recall may thus be a reflection of differences among individuals in the utilization of repression. Other correlates of repression have been derived from the research in cognitive controls (Gardner, Holzman, Klein, Linton, & Spence, 1959) which has demonstrated that the leveling-sharpening tendency is related to the operation of the ego defenses, specifically repression. Holzman and Gardner (1959) compared scores on the Schematizing test (described below) with ratings of repression, based on Rorschach test indicators, e.g., few responses, vagueness, emphasis on color. All extreme repressors were also extreme levelers although the converse was not found.

The studies of Schonbar (1959), and Holzman and Gardner (1959) suggest that dream recall, repression and the leveling-sharpening continuum are associated. It is the hypothesis of the present study that subjects who recall few if any dreams, i.e., who use repression as a major defense, will show the leveling tendency. It is further hypothesized that subjects who recall many dreams, i.e., do not rely upon repression as their major defense, will show the sharpening tendency.

The intention here is to relate empirically two conceptual frames of reference, the psychoanalytic concept of repression and the cognitive controls concept of leveling-sharpening, with respect to the recall of dreams.

#### METHOD

The subjects were students in two introductory psychology classes at a municipal college. The tests utilized were administered by one of the authors. Since he was also the instructor of the

<sup>1</sup> The authors are indebted to Lloyd Silverman for his suggestions in formulating this study and to Edward Bryant for his assistance with the apparatus.

Frank M. Lachmann and Norman S. Handelman are now associated with the Postgraduate Center for Psychotherapy, New York. Benjamin Lapkin is now at Hillside Hospital, New York.



classes, it was felt that at the time of testing sufficient rapport had been established for cooperation. Material distributed was numbered to permit subsequent collation. The subjects were assured of anonymity and that nonparticipation would in no way affect their class standing. The two classes totaled 52 students, 11 males and 41 females. The mean age of the group was 19.7 years with a range from 17.5 to 42.7 years. Six of the group were married. No student had received psychotherapy or psychoanalysis.

### Recording of Dreams

Following Schonbar (1959), booklets were prepared in which the subjects were to record daily their dream activity for 14 consecutive days. Space was provided to record the content of the dreams or to note that no dream was recalled. The distribution of booklets to the subjects was accompanied by instructions identical to those published by Schonbar. While she collected the dream reports daily for 4 weeks, the present procedure involved a 2-week period, at the end of which the entire booklet was collected. On alternate days, the subjects were reminded to continue recording in the booklets. The measure of dream recall was the number of days in which subjects recorded at least one dream. The sample was divided into quartiles based upon this measure, the upper quartile being the high recall group and the lowest quartile, the low recall group.

### Schematizing Test

One week after the distribution of the dream report booklets, the Schematizing test was administered to each class as a group. Subjects were told that they were taking a test of perceptual accuracy and no association was made with the dream study.

A detailed description of the apparatus and administration of this test appears in Gardner et al. (1959). The test required subjects to estimate sizes of squares projected individually onto a screen. There were 10 series of 15 squares each, the first 5 squares of each series in an ascending order and the subsequent 10 squares of the same sizes began by a random order. Each following series began by dropping the smallest square of the preceding series and ascending one step. The smallest square in the first series was 1.2 inches and the largest square in the tenth series was 13.7 inches. The 150 slides were presented automatically at a uniform exposure of 3 seconds, with an 8-second interval between exposures. An electromagnetic shutter, operated by Hunter timers, controlled the presentations of an automatic load 35-millimeter slide projector.

As described by Holzman and Gardner (1959), the Schematizing test was scored for both ranking accuracy and degree of lag. A composite score was obtained for each subject based upon this double criterion. Subjects were classified as levelers if they

scored in the upper third and sharpeners if they scored in the lower third of the distribution.

### Subjects

Both parts of the study were completed by 38 subjects, 8 males and 30 females. The remainder either failed to return the dream report booklets, returned the booklets with an indication that they did not participate, or were absent when the Schematizing test was administered. In view of sex differences in the Schematizing test noted by Gardner et al. (1959), and in consideration of the small number of men in the sample, only females were included in the following analysis of data. The final group evaluated consisted of 30 women, ranging in age from 17.7 to 42.7 years, with a mean age of 19.9. Five were married.

### RESULTS

Tabulation of the dream recall reports indicated a range of days on which dreams were recorded from 0 to 10 with a median of 4 days. The upper quartile consisted of 9 subjects who reported dreams on 6 days or more. The lowest quartile of 7 subjects who reported the lowest quartile of 7 subjects who reported dreams on no more than 1 day and the middle group consisted of 14 subjects who reported dreams on 2-5 days. A classification of subjects on their Schematizing test scores yielded 11 levelers, 10 sharpeners, and a midgroup of 9 subjects. Table 1 presents the number of levelers and sharpeners in each of the dream recall categories.

An analysis of variance was computed, comparing the subjects in the high, medium, and low dream recall groups with respect to their Schematizing test scores. Table 2 presents this analysis.

An  $F$  value of 3.31 was obtained which approximates the  $F$  value of 3.36 required for significance at the .05 level. In view of this finding a  $t$  test was computed between the two groups which showed the greatest difference between their Schematizing test scores: the medium and high dream recall groups ( $t = 2.67$ ,  $p$  between .01 and .02). A  $t$  test was also computed be-

TABLE 1  
NUMBER OF LEVELERS AND SHARPENERS  
IN THE DREAM RECALL GROUPS

Schematizing test category	Dream recall groups		
	High	Medium	Low
Levelers	1	7	3
Sharpeners	5	2	3
Unclassified	3	5	1



TABLE 2  
ANALYSIS OF VARIANCE OF SCHEMATIZING  
TEST SCORES

Source of variation	SS	df	MS	F
Between groups	1079.33	2	539.67	3.31
Within groups	4406.64	27	163.21	
Total	5485.97	29		

tween the medium and low dream recall and a  $t$  value of 1.15, below an acceptable level of significance, was obtained. This indicates that high dream recallers tend to score toward the sharpening end of the continuum. Subjects in the medium dream recall group tend to score toward the leveling end. No definite results were obtained with regard to the low dream recall group. Inspection of the data of this latter group cited in Table 1 reveals the presence of three levelers and three sharpeners.

To evaluate the overall relationship between number of dreams and the leveling-sharpening continuum a Pearson product-moment correlation coefficient was computed. A nonsignificant correlation of .20 was obtained. Thus though high dream recall is associated with the sharpening tendency and medium dream recall with leveling, it cannot be substantiated at this point that low dream recall fits into this continuum.

### DISCUSSION

The findings of this study point to a difference between the medium and high dream recall groups with respect to cognitive controls. High dream recall is associated with the cognitive control of sharpening, thus supporting one of the hypotheses of this study. It had been hypothesized initially that the converse would hold also, namely, that the low dream recallers would show the leveling tendency. However, the results do not support this contention, indicating instead that low dream recallers actually are at either extreme of the schematizing continuum. It thus appears that the repression-leveling relationship is not supported. This is contrary to the findings of Holzman and Gardner (1959) and Gardner et al. (1959) who reported a significant association along these lines. To understand these divergent findings the following possibilities are presented. First, different criterion measures were used: Holzman and Gardner used Rorschach protocols to measure reliance upon repression

while the present investigators used the amount of dream recall. Furthermore, the absence of normative criteria for leveling and sharpening on the Schematizing test makes interstudy comparisons difficult. Further investigation is needed to provide normative data in this area.

### SUMMARY AND CONCLUSIONS

The present study investigated the relationship between the recall of dreams and the cognitive controls continuum of leveling-sharpening. Thirty college students reported dream activity daily for a 14-day period and the amount of dream recall was related to their scores on the Schematizing test. The results confirmed the hypothesis that subjects who recall many dreams, i.e., who do not rely primarily upon repression, would show the sharpening tendency. A second hypothesis, postulating a relationship between low dream recall and the leveling tendency, was not supported.

It appears that the psychoanalytic concept of repression can be associated with the theory of cognitive controls, in that there is a statistical relationship between repression as measured by the extent of dream recall and the leveling-sharpening tendency as measured by the Schematizing test.

### REFERENCES

- DEMENT, W., & KLEITMAN, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *J. exp. Psychol.*, 1957, 53, 339-346.
- FREUD, S. *The interpretation of dreams*. (Originally published 1900) New York: Basic Books, 1959.
- GARDNER, R., HOLZMAN, P. S., KLEIN, G. S., LINTON, HARRIET, & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychol. Issues*, 1959, 1, No. 4.
- GOODENOUGH, D. R., SHAPIRO, A., HOLDEN, M., & STEINSCHRIEBER, L. A comparison of "dreamers" and "nondreamers": Eye movements, electroencephalograms, and the recall of dreams. *J. abnorm. soc. Psychol.*, 1959, 59, 295-302.
- HOLZMAN, P., & GARDNER, R. Leveling and repression. *J. abnorm. soc. Psychol.*, 1959, 59, 151-155.
- KLEITMAN, N. Patterns of dreaming. *Scient. American*, 1960, 203, 82-88.
- RAMSEY, G. V. Studies of dreaming. *Psychol. Bull.*, 1953, 50, 432-455.
- SCHONBAR, ROSALEA A. Some manifest characteristics of recallers and nonrecallers of dreams. *J. consult. Psychol.*, 1959, 23, 414-418.

(Received January 30, 1961)

# Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

## CONSULTING EDITORS

JOSEPH B. ADELSON  
*University of Michigan*

GORDON W. ALLPORT  
*Harvard University*

NORMAN H. ANDERSON  
*University of California, Los Angeles*

LEONARD BERKOWITZ  
*University of Wisconsin*

JACK W. BREHM  
*Duke University*

URIE BRONFENBRENNER  
*Cornell University*

DONALD R. BROWN  
*Bryn Mawr College*

DONALD T. CAMPBELL  
*Northwestern University*

LAUNOR F. CARTER  
*System Development Corporation*

DWIGHT W. CHAPMAN, JR.  
*Vassar College*

MORTON DEUTSCH  
*Bell Telephone Laboratories*

SEYMOUR FESHBACH  
*University of Pennsylvania*

DONALD W. FISKE  
*University of Chicago*

NATHANIEL L. GAGE  
*University of Illinois*

NICHOLAS HOBBS  
*Peabody College for Teachers*

MURRAY HORWITZ  
*New York University*

IRVING L. JANIS  
*Yale University*

HAROLD H. KELLEY  
*University of California, Los Angeles*

HERBERT C. KELMAN  
*Harvard University*

DANIEL J. LEVINSON  
*Harvard Medical School*

DAVID C. MCCLELLAND  
*Harvard University*

ROBERT B. MACLEOD  
*Cornell University*

HAROLD L. RAUSH  
*National Institute of Mental Health*

NEVITT SANFORD  
*Stanford University*

IRVING SARNOFF  
*Western Reserve University*

WILLIAM A. SCOTT  
*University of Colorado*

JOSEPH VEROFF  
*University of Michigan*

JUSTIN L. WEISS  
*Massachusetts Mental Health Center*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is \$10.00, or \$20.00 annually (foreign, \$20.50). Single issues are \$2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ  
University of Michigan  
Box 1248  
Ann Arbor, Michigan

---

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR  
Acting Production Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 Sixteenth Street, N.W.  
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
Prince and Lemon Sts., Lancaster, Pa.  
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Lancaster, Pa., and at additional mailing offices

Copyright © 1962 by the American Psychological Association, Inc.



## THE RESTRICTING EFFECTS OF AWARENESS: A PARADOX AND AN EXPLANATION<sup>1</sup>

DONALD P. SPENCE AND BERT HOLLAND

*Research Center for Mental Health, New York University*

An increasing body of data in recent years has shown that a very brief or faint perceptual stimulus may have an effect on behavior, but the nature of this effect and ways to account for it are still under discussion. The effect is generally slight (as in studies by Klein, Spence, Holt, & Gourevitch, 1958; Paul & Fisher, 1959; Smith, Spence, & Klein, 1959) and in some cases open to the suspicion that partial awareness of the stimulus could account for the response. Such critics as Eriksen have argued that some aspects of the stimulus were consciously perceived but un verbalized; since we depend on a verbal response ("I didn't see anything") to measure awareness, we are misled into thinking that no conscious cues are being received. Because the insensitive verbal indicator is used to determine threshold, it leads to the false conclusion that a stimulus is subliminal. In reality, says Eriksen (1960), the subject is receiving conscious information which he cannot report. Because the cues are fragmentary, the response is usually inexact, in proportion to the indistinct quality of the stimulus.

The study by Smith et al. (1959) can be used to illustrate this argument. Two words, HAPPY and ANGRY, were briefly exposed and immediately followed by a consciously seen, ambiguous face. Descriptions of the face appeared to be influenced by the unseen verbal stimulus but the effects were indirect, the descriptions rarely containing the words HAPPY or ANGRY. Since a later check showed that many of the exposures of the verbal stimuli were at or above recognition threshold, some

partial information was possibly available, and it could be argued that the subject elaborated this fragment of information into a general description of the face. The experimental effect was presumably slight because the stimuli were exposed only briefly (though repeatedly).

If a brief and impoverished stimulus produces only a primitive and allusive response, it is not surprising that further decreasing the amount of input will lead to a gross autonomic response. Such a finding was reported by Lazarus and McCleary (1951) in a study which showed that PGR measures were sensitive to the stimulus in the absence of verbal report. If the amount of information is increased, on the other hand, the amount of effect upon subsequent responses should increase, and this assumption is supported by Bressler (1931) who showed that the extent of the effect varied directly with the intensity of the stimulus. In short, one can postulate that increments in stimulus intensity lead to increments in adequacy of response and we will label this the intensity argument. It makes two general assumptions:

Assumption 1. A response based on partial cues is on the same continuum with a response to a fully developed stimulus, but the former is a paler and less precise copy of the latter. The notion of one continuum has been discussed recently by Weiner and Schiller (1960).

Assumption 2. Awareness of the stimulus can be used to gauge the amount of information being registered, so that when awareness is minimal (as in the subliminal studies) only partial cues are being received and only a primitive and inarticulate response can result. Increased awareness is a sign that more information is being presented which improves the "fit" between stimulus and response.

The intensity argument accounts for the main effects of the subliminal studies with-

<sup>1</sup> The study was carried out as part of a research project supported by grants from the United States Public Health Service.

We are indebted to Charles Cofer, Robert R. Holt, George S. Klein, Fred Pine, and I. H. Paul for their many helpful suggestions, and are especially grateful to Earl Reynolds, formerly associated with the School of Commerce, New York University, for making available his classes and classrooms.

out postulating a process below awareness, and in the minds of some (Eriksen, 1960; Weiner & Schiller, 1960) this is a victory for parsimony. But one finding, which has appeared repeatedly, does not fit this argument—the finding that the effect of an impoverished stimulus varies inversely with its intensity. This finding first appeared in the study by Smith et al. (1959) in which the verbal stimuli were exposed at progressively longer exposures, descriptions being obtained after each exposure. Effects of these words on the ambiguous face were much stronger and more systematic at low levels of exposure than at high levels. An analogous finding was reported by Paul and Fisher (1959): their stimulus had a greater effect on images when the level of exposure was considerably below the subject's threshold than when it was just below the subject's threshold. Eagle (1959) reports that subjects most influenced by subliminal stimuli had the highest thresholds. Lapkin and Lippmann (1959), report a correlation of .57 ( $p < .05$ , one-tailed test) between the degree to which the stimulus exposure was below threshold, and the magnitude of its effect on the response measure. The implication is that as information (exposure level) is decreased the effect increases, in contradiction to the first assumption of the intensity argument; and in contradiction to the second assumption, we find that awareness cannot be used to predict the amount of effect.

These findings seemed to embarrass the intensity argument, which uses awareness as an *indicator* of registration, and they provided the impetus for another approach. First advanced by Freud (1953) and later elaborated by Klein et al. (1958) and Fisher (1957), this approach makes four important assumptions: registration of the stimulus is independent of awareness, with awareness and registration conceptualized as separate aspects of cognitive activity; degree of awareness of a stimulus at the moment of input cannot be used to predict the extent of its effect on a response; the effects of a faint and weak stimulus are not necessarily weaker than those of a strong and clear stimulus; and reduction of awareness changes the patterning of cognitive activity.

Where the intensity argument implies that responses diminish to zero as the threshold is approached, the second argument states that a new class of responses is produced as we go below threshold. (We are not talking here of a physiological threshold. Obviously there is a lower limit below which no responses are possible.) The germ of this argument was set forth in the *Interpretation of Dreams* where Freud (1953) noted that thinking below awareness tends to fan out over a large network of traces.

The energy attaching to the train of thought is diffused along all the associative paths that radiate from it; this energy sets the whole network of thoughts in a state of excitation which lasts for a certain time and then dies away . . ." (p. 594).

Freud also assumed that the associations in this network were logically organized and followed the rules of the secondary process. Putting these assumptions together we find that reduction of awareness should not necessarily lead to a more primitive response but to a greater spread of activity to other logically related responses. One idea can spread more freely throughout the system than is possible in conscious thinking where it is usually restricted to one train of thought; one percept can make contact with a wider range of traces. Support for this assumption is provided by experiments by Lacey and Smith (1954), Spence (1961), and Fox (1960), all of which show that, when not in awareness a verbal stimulus may generalize to a broad range of meanings. The generalization is often greater than is found with a conscious stimulus, and this finding suggests that when awareness is reduced, certain restrictions on thought are removed; more rather than fewer associative pathways become available and thought is freer to range over a greater number of associates.

Suppose a test word is exposed below threshold and followed by a list composed of words that are either related or unrelated to the test word. Subjects are then asked to recall the word list. Suppose another group of subjects undergoes the same procedure except that the test word is presented above awareness. The intensity hypothesis would predict that the subliminal group, receiving only a briefly exposed stimulus, could respond to it only on



the basis of partial cues—an isolated letter, for example—and would therefore tend to recall list words composed of the same, or similar, letters. These effects would vary with awareness: subjects who could barely discriminate the test word would get less information and would show a more random recall than subjects with a greater number of correct discriminations. If the latter group perceived enough information they might apprehend the meaning of the stimulus and recall related words. Thus the number of correct discriminations should be positively correlated with recall of related words. The supraliminal group, fully aware of the test word, also should respond in terms of meaning and, because it gets more information than the first group, should remember more words on the list which were meaningfully related to the stimulus.

But from our alternative hypothesis we would draw opposite predictions. We would assume that the subliminal stimulus would make contact with a wide range of traces and facilitate the recall of many associated words. Because thinking tends to fan out when it goes on outside of awareness (awareness, we propose, has a restricting effect), the spread of effect will be *inversely* correlated with amount of the stimulus in awareness. Thus subjects with few correct discriminations of the test word would be expected to recall more of the meaningfully related list words than would subjects with a high number of discriminations, and subjects in the Subliminal group would recall more such words than subjects in the Supraliminal group. The influence of partial cues would be relatively slight, subordinated to the influence of meaning, and not perfectly related to amount of awareness.

#### METHOD

If the meaning of a word registers outside of awareness and fans out along associative pathways, it would be likely to influence other associates which can be determined by word association norms. From the Minnesota norms recently collected by Russell and Jenkins (1954), we selected the word CHEESE as our stimulus and 10 associates to CHEESE which range from popular (MOUSE) to rare (BRICK). We chose words which appeared to be minimally interrelated to prevent the recall of one word from triggering off the others; then the number of words re-

TABLE 1  
SETS OF STIMULUS AND OF CONTROL WORDS

Minnesota frequency <sup>a</sup>	CHEESE associates	T-L frequency <sup>b</sup>		Control words
94	MOUSE	34	32	hedge
82	BREAD	A	A	flag
32	COTTAGE	46	46	mirror
21	SMELL	A	A	gift
19	MOON	AA	AA	best
18	GREEN	AA	AA	chair
7	COW	A	A	sand
6	SOUP	15	16	frail
2	CAVE	33	34	arch
1	BRICK	49	48	trunk

<sup>a</sup> Responses per 1,008 subjects.

<sup>b</sup> Occurrences per million words. A = 50 per million; AA = 100 or more per million. Words were presented in the following order: street/bird/dime/flag/cow/arch/BREAD/CAVE/chair; MOON, best/COTTAGE/gift/frail/GREEN/sand/BRICK/hedge SMELL/MOUSE/mirror/SOUP/trunk/chin/number/joke. The first and last sets of three words are to control for primacy and recency effects and do not enter into the analysis of the findings.

called could be used as a measure of size of effect. We also selected 10 control words which were not listed as associates to CHEESE, but which matched the CHEESE associates in Thorndike-Lorge (1944), frequency, number of letters, and part of speech. The two sets of words are presented in Table 1.<sup>2</sup>

Further support for our assumption that CHEESE associates and control words are equally balanced for recall can be noted in the finding that of 35 subjects so far tested in a blank condition (19 from this study plus 16 from another study), the mean recall of CHEESE associates (2.83) was not significantly different from the mean recall of control words (2.57).

If meaning is an important factor in recall, more CHEESE associates should be recalled than control words. If, on the other hand, fragments of a word register instead of its meaning, recall of structurally similar words should exceed recall of structurally different words. To test this hypothesis, the stimulus word CHEESE was divided into the two-letter combinations of ch, he, and es and the control words divided into a group which contained these letter-pairs (chair, arch, hedge, and best) and those which did not (gift, flag, mirror, and trunk). If partial information influenced recall, the first group would be recalled better than the second group (none of which contained a single letter in common with CHEESE).

<sup>2</sup> Subsequent data have supported our assumption that the words are minimally interrelated. As part of another study, 70 subjects gave word associations to each of the 26 words on the recall list. One CHEESE associate produced another CHEESE associate as a response in only 5 cases out of 700; one control word produced another control word as a response in only 2 cases out of 700. The interitem associative strength, to use an expression coined by Deese (1959), is, therefore, essentially zero and there is little reason to assume that stimulus words cluster any more readily in recall than control words.



### Procedure

Our design called for two groups equated for associative tendencies, and to obtain the relevant data we presented a large class with a list of 25 words taken from the Minnesota norms. Jenkins and Russell (1958) report that the number of popular associations to these 25 words is highly correlated with the number of populars given to the entire list of 100 words. Associative preference (commonality) was determined for 48 subjects and they were divided into two groups of equivalent mean commonality. One group was assigned to the Subliminal condition, and the other to a Blank condition.

*Blank and Subliminal conditions.* Two days after the selection procedure the experimental procedure was carried out. The Subliminal group was assigned to seats facing the front of the room and arranged in five equal rows of equal mean commonality. The front row was situated 21 feet from the screen; at that distance, the visual angle of the stimulus was  $7^\circ$ . The back row was 33 feet from the screen, with a visual angle of  $4.3^\circ$ . Subjects in the Blank group were seated around the edge of the room and faced away from the screen.

The experiment consisted of the exposure of the test word (CHEESE), a single reading of the word list, recall of the word list, and a measure of threshold. Subjects in the Blank group were told to close their eyes, while subjects in the Subliminal group were asked to look at a fixation point on the screen, and were told that something would be flashed in the neighborhood of that point. They were to write down what they thought it was, and how many times it flashed. The projected stimulus extended 1 foot 3 inches on either side of the fixation point. It was flashed five times at  $1/150$ th of a second from a slide projector with an Ilex shutter. Room illumination was maximized, and the shutter diaphragm was set at its smallest opening. The combination of high room illumination, low stimulus intensity, and brief exposure worked effectively to prevent conscious recognition. No subject reported seeing anything, and the general reaction of the group was that nothing had been flashed. Estimates of the number of exposures varied from 4 to 5, but this seems to have been based on the number of shutter clicks heard by the subjects.

The Blank group was then asked to open their eyes, face front, and both groups were given the following instructions:

We're going to read a list of 26 words. Later on we will ask you to reproduce them from memory. This is not a test, for it is quite impossible to remember all the words. Just relax and listen closely while we read the list. We will read it through once.

The list was then read by an instructor who had no knowledge of the purpose of the experiment. Three "padding" words were read first to absorb the effect of primacy; then the 10 CHEESE associates

and 10 control words were presented in scrambled order; then three more "padding" words were added to control for recency. The complete list is given in Table 1.

Immediately after the list was read, subjects were told to write whatever words they could remember, in whatever order, and to continue writing for as long as they wished. They were encouraged to guess.

*Threshold.* To check on the subject's degree of awareness of the test word, we followed a suggestion advanced by Goldiamond (1958) and presented the stimulus and a blank slide in a series of 20 exposures in a fixed random order. Each slide was presented under the same condition of illumination and exposure as were used in the main part of the experiment. If any subjects had correctly identified CHEESE they would have been eliminated from the group. Subjects were given these instructions:

Watch the screen once again, at the dot. Sometimes we will show you something very faintly and hard to see. Other times we will show you nothing. After each flash, write something or nothing. If you saw anything, please describe it at the bottom of the page.

The number of correct discriminations provided a measure of the relative exposure level of the stimulus.

*Supraliminal condition.* Associative commonality was determined for a separate class using the procedure described above; 2 days later subjects were arranged in front of the screen in such a way that each row had the same average commonality, and the distance of each row from the screen matched the distance for the Subliminal group.

Subjects were told to expect something to be flashed on the screen. They were told to watch the fixation point and write down what they saw, and how many times it flashed. CHEESE was exposed 5 times in exposures of 2 seconds each. Its projected size and placement on the screen were the same as in the previous Subliminal session. The group was then read the same preparatory instructions used in the first experiment; the standard list was read by the same instructor, and subjects wrote down their recall. All subjects perceived the stimulus CHEESE, and all but one counted five flashes (he counted four).

### Subjects

The members of all three groups (Subliminal, Supraliminal, and Blank) were taken from the undergraduate classes of the School of Commerce, New York University. Of the 65 subjects, 59 were male. Twenty-nine subjects composed the Subliminal group, 19 the Blank group, and 17 the Supraliminal group. The Subliminal group was made purposely larger than the rest in order to better investigate the phenomenon. All three groups were taught by the same instructor and the experiments took place during regular class sessions. All subjects and the instructor were naive as to the purpose of the experiment; no subjects were majoring in psychology.

## RESULTS

*Effect of Subliminal Stimulus on Recall*

The stimulus CHEESE was not consciously perceived by any subjects in the Subliminal group. Six of the 29 subjects did detect the presence of "something" in the discrimination series significantly better than chance ( $p < .05$ , one-tailed test), and they were eliminated from the main analysis in the interest of caution. The recall of the remaining 23 subjects was analyzed for effects of structure and meaning of the subliminal stimulus.

*Structure.* If the structure of the stimulus CHEESE influenced recall, more structurally related control words (*chair, arch, hedge, and best*) should be recalled than dissimilar control words (*gift, flag, mirror, and trunk*). There is a trend in this direction but it is not significant ( $.10 > p > .05$ , two-tailed, Wilcoxon test). However, if we add the six subjects who were able to discriminate the stimulus better than chance, the difference reaches significance ( $p < .01$ , two-tailed, Wilcoxon test) because these six subjects were responding primarily on the basis of structural cues. Their recall of the structurally similar words ranged from 1 to 3, with a mean of 2.2; their recall of the dissimilar words ranged from 0 to 2, with a mean of .33. When partial cues are available to the Subliminal group they facilitated recall of structurally similar words, as the intensity hypothesis would predict; *but this is only true at the level of partial awareness.* Below this level these cues have no significant effect, and their effect does not vary with intensity or awareness of the stimulus. Using only the 23 "chance discriminating" subjects, we computed each subject's recall of structurally similar minus dissimilar words; this net recall is correlated  $-.13$  ( $ns$ ) with distance from the screen—a measure of size and intensity of the stimulus—and  $-.28$  ( $ns$ ) with number of discriminations—a measure of awareness. Thus the intensity argument is not supported at subliminal levels of exposure.

For the Blank group, the difference in recall of the two groups of structurally related words was not significant.

*Meaning.* If the meaning of the CHEESE

TABLE 2

MEAN RECALL OF CHEESE ASSOCIATES AND CONTROL WORDS UNDER THREE CONDITIONS

Condition	N	CHEESE associates	Control words	Difference*
Subliminal	23	4.09	1.91	2.18*
Blank	19	2.89	2.26	.63
Supraliminal	17	3.00	2.41	.59

\* Differential recall is significantly higher in Subliminal than in Supraliminal or Blank conditions ( $.01 > p > .001$ , Kruskal-Wallis test). Differential recall is significantly higher in Subliminal than in Blank ( $p < .02$ , two-tailed, Mann-Whitney  $U$  test); significantly higher in Subliminal than in Supraliminal ( $p < .02$ , two-tailed, Mann-Whitney  $U$  test); and not significantly different between Blank and Supraliminal.

\* Significant at .001 level. Significantly more CHEESE associates than control words are recalled in the Subliminal condition (two-tailed, Wilcoxon test). Recall between CHEESE and control words is not significantly different in the other two conditions.

stimulus influenced recall, more CHEESE associates should be recalled than nonrelated control words. Nineteen of the 23 subjects in the Subliminal group recalled more CHEESE associates than control words ( $p < .001$ , two-tailed, Wilcoxon test), a strong indication that meaning had an effect.<sup>3</sup> Only 9 of the 19 subjects in the Blank group recalled more CHEESE associates ( $ns$ ). These findings strongly suggest that the meaning of a stimulus that lies outside the range of correct discrimination can influence the recall of related words. The strong subliminal effect is presented in Table 2.

The associates recalled by the 23 subject Subliminal group cover the full range of response frequencies. We computed the percentage of subjects in the Subliminal group who recalled each CHEESE associate and the rank-order correlation between these frequencies and Minnesota frequencies is .42 ( $ns$ ). Since some of the recall probably depended on uneven stresses in pronunciation, uneven pausing, and other characteristics of the manner in which the list was read, we computed comparable percentages for subjects in the Blank group, and subtracted one set of frequencies from the other. Since both groups listened to the same list, any difference in re-

<sup>3</sup> A replication has recently been completed by the senior author. Nineteen subjects received the subliminal stimulus and 16 subjects served as controls, using the same procedures described above. Recall of CHEESE associates minus control words was significantly better in the Subliminal than in the Blank group ( $p < .06$ , one-tailed test).

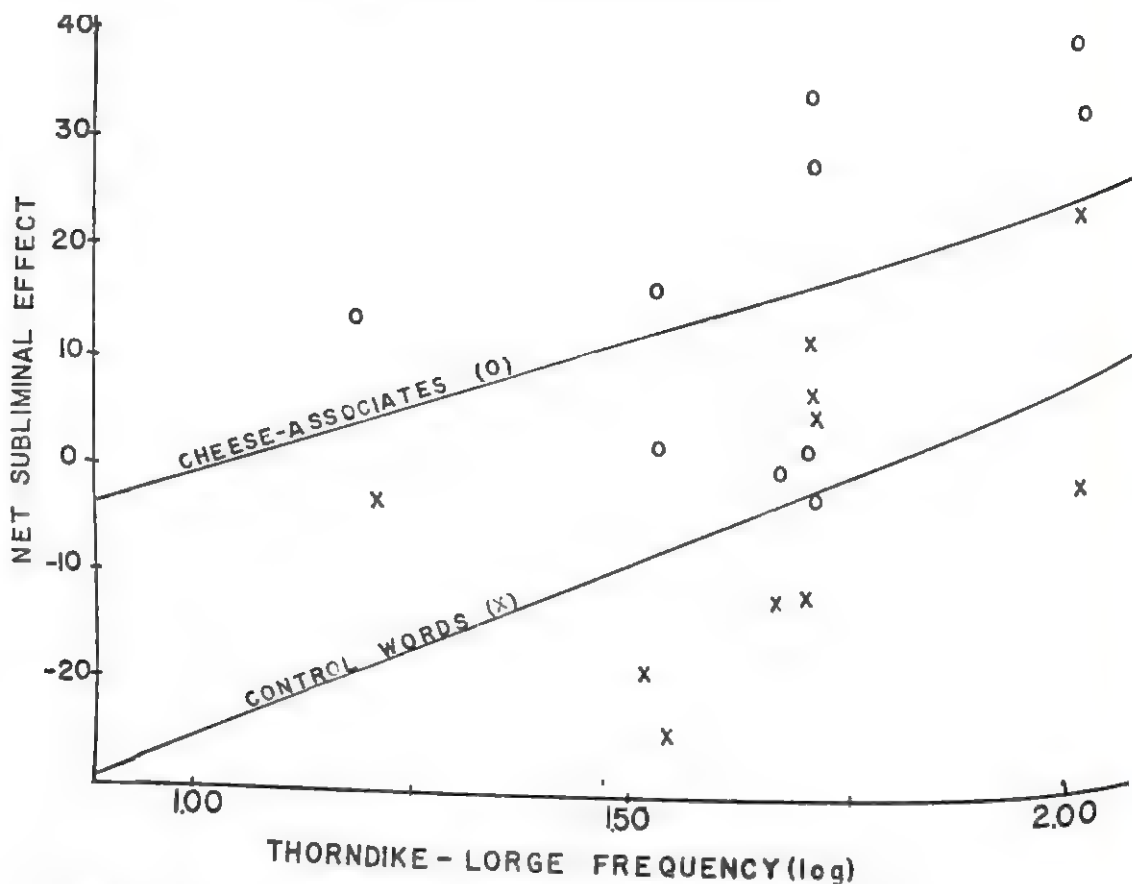


FIG. 1. Net subliminal effect (percentage of recall in the Subliminal group minus percentage of recall in the Blank group) as a function of Thorndike-Lorge frequency. (Regression lines as indicated.)

call was presumably due to the subliminal stimulus and was free of "list effect." Although these differences are likewise not correlated with Minnesota frequency ( $\rho = -.10$ ), they are strongly related to Thorndike-Lorge word frequency ( $\rho = .63$ ,  $p < .05$ , two-tailed test).<sup>4</sup> Within the group of CHEESE associates, the extent of recall is determined more by the general popularity of each word than by its popularity as an associate to CHEESE.

A correlation with word frequency is also reflected in the net recall of control words ( $\rho = .78$ ,  $p < .02$ ). Apparently the frequency of occurrence determines the readiness of recall; words that are infrequently used in written English are unfamiliar and less likely to be recalled. If a word was meaningfully related to CHEESE, it received an additional increment which seems to be more or

less independent of its association strength (see Figure 1 in which the gradient of recall for the CHEESE associates is roughly parallel, but shifted above, the gradient for the control words). The likelihood of recall for a particular word appears to have been the sum of its initial habit strength plus a more-or-less uniform subliminal reinforcement.

However, when word frequency is partialled out, we do find an interesting relation to associative strength. Amount of net recall of the five most popular CHEESE associates (A or AA words on Table 1) is correlated  $-.90$  ( $p < .10$ , two-tailed test, rank-order correlation) with Minnesota frequency which shows that the subliminal stimulus has a stronger effect on its weaker associates, other things being equal.

Since recall was influenced by meaning, we next asked whether it was influenced by the subject's associative preference (commonality); it was not. Although subjects with high

<sup>4</sup> We are indebted to George S. Klein for suggesting the existence of this relationship.



commonality tend, by definition, to give popular responses on an association test, they did not tend to recall popular CHEESE associates. The average association strength of the associates recalled by subjects is not correlated with commonality ( $\rho = -.06$ ). These results show that the subliminal stimulus reinforces a broad network of associates which are not concentrated around the most preferred conscious response and (to judge by the preceding paragraph) tend to be distantly related to the stimulus. Apparently the subliminal stimulus follows laws that are independent of those which determine conscious associations.

### *Influence of Awareness on Recall*

Recall of CHEESE associates in the Supraliminal condition was not significantly greater than control words (see Table 2) which indicates that exposure of an organizing word does not necessarily lead to an increase in its associates. This finding parallels that of Deese (1959) who found that recall of associates was not increased when the list was preceded by the stimulus for those associates. Moreover, we find that differential recall of CHEESE associates is higher in the Subliminal than in either the Blank and Supraliminal conditions, and differences between conditions are highly significant ( $.01 > p > .001$ , Kruskal-Wallis one-way analysis of variance). Table 2 clearly shows that the significance stems from the high difference in the Subliminal condition, and subsequent Mann-Whitney tests reveal that differential recall is significantly greater in Subliminal than in Supraliminal ( $p < .02$ , two-tailed test) and greater in Subliminal than in Blank ( $p < .02$ , two-tailed test). The difference between Blank and Supraliminal is clearly not significant. Apparently the greatest effect on recall of CHEESE associates is produced by the subliminal stimulus; comparing subliminal and supraliminal effects, we find that awareness of the stimulus apparently restricts its effect on recall. If this is generally true, those subjects in the Subliminal group who were maximally aware of the stimulus should recall fewer CHEESE associates than those subjects who had minimal awareness of the stimulus. To test this assumption, we re-

placed the six subjects who had been dropped earlier from the Subliminal group and computed the correlation between number of discriminations, which ranged from 8 to 19 out of 20, and relative recall (CHEESE associates minus control words). The rank-order correlation is  $-.54$  ( $p < .02$ , two-tailed test); poorer discrimination is accompanied by greater net recall of words which are meaningfully related to CHEESE (see Figure 2). As discrimination improves the effect falls off.

In view of the greater effect of partial cues among high discriminating subjects noted earlier, and the greater effect of meaning among the randomly discriminating subjects, we made a direct comparison of the two factors of structure and meaning. We compared recall of arch, chair, best, and hedge, the four structurally similar control words, with four CHEESE associates of equal word frequency—CAVE, GREEN, MOON, and MOUSE. Two of these words contain pairs of letters which are identical to a pair of letters in the test stimulus; one has similar letters not contained in pairs; and the fourth has no similar letters. We found that subjects at the low end of the discrimination continuum recalled more CHEESE associates than control words; that subjects at the high end recalled more control words than CHEESE associates; and that, overall, there is a correlation of  $.57$  ( $p < .02$ ,

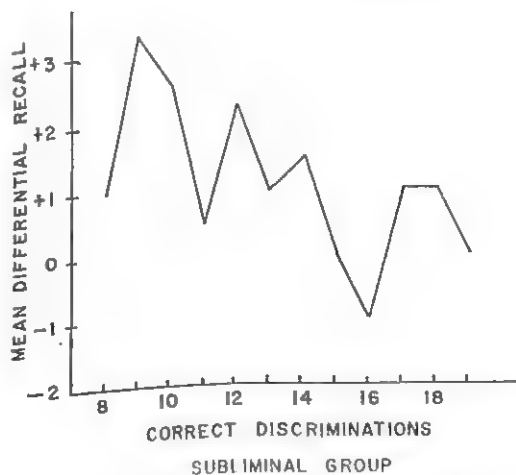


FIG. 2. Mean differential recall (CHEESE associates minus control words) as a function of correct discrimination in the subliminal condition.

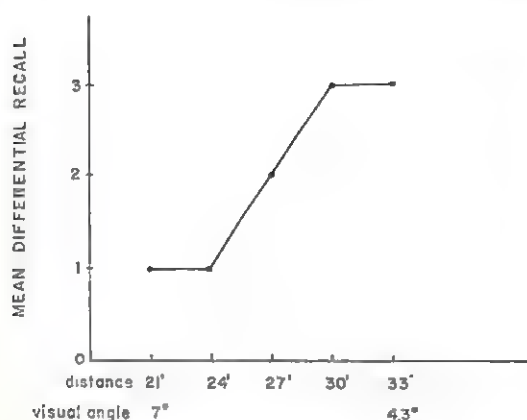


FIG. 3. Mean differential recall (CHEESE associates minus control words) as a function of distance and visual angle in the subliminal condition.

two-tailed test) between number of discriminations and the preference for control words. Evidently structural cues gave way to meaning at low levels of awareness, but began to organize recall at a high level of awareness. These findings suggest that partial cues, as they became more available, tended to organize the recall in terms of structural elements and thus reduced the recall of meaning associates (which have less structural correspondence to the test stimulus).<sup>5</sup>

Another kind of limiting effect can be seen when we study the intrusions—words which were “remembered” but were not on the original list. The supraliminal stimulus tended to direct attention toward food associates to CHEESE (sweet, taste, meat, etc.) as shown by the fact that the number of CHEESE related intrusions is significantly higher in the Supraliminal group ( $p < .001$ ) than in the Subliminal group, even though the average number of all intrusions is the same in both conditions.

Putting these findings together, we find that awareness of the stimulus tended to reinforce a particular area of meaning and restricted

<sup>5</sup> This finding is not simply a restatement of the general trend expressed in Figure 2 for meaning-effect to decrease as awareness increases. A similar comparison of two other sets of words (BREAD, SMELL, COTTAGE, and BRICK vs. flag, gift, mirror, and trunk) does not show any correlation with awareness ( $\rho = .25$ ); i.e., the high discriminating subjects are not less sensitive to meaning when the equivalent control words (flag et al.) have no structural similarity to the test stimulus.

sensitivity to other connotations. Overemphasis on one category of meaning markedly reduced the range of words reinforced by the perceived word CHEESE and led to the paradoxical finding that a supraliminal stimulus had a more restricting effect on recall than a subliminal stimulus.

### *Influence of Intensity and Size on Recall*

Our subjects in the Subliminal condition were arranged in five rows of varying degrees of distance from the screen, and by taking advantage of the fact that intensity and size vary with distance, we had a chance to measure how these variables influence effect. Recall of CHEESE associates minus recall of control words was positively related to distance from the screen ( $\rho = .63$ ,  $p < .02$ , two-tailed test; see Figure 3). At the maximum distance, the visual angle ( $4.3^\circ$ ) is only slightly greater than the foveal angle and thus the most sensitive part of the retina was being stimulated by the test word.<sup>6</sup> At shorter distances less of the word was impinging on the fovea, in which case less of its meaning would be communicated and we would expect that only single letters of the test word would have an effect on recall. Using the same two groups of words mentioned earlier, we found that subjects in the front rows recalled more structurally similar control words (arch, chair, best, and hedge) than the more dissimilar CHEESE associates, (CAVE, GREEN, MOON, and MOUSE); that subjects in the back rows recalled more dissimilar CHEESE associates than similar control words; and that, overall, preference for meaning vs. structure as an organizing principle was correlated .64 with distance ( $p < .02$ , two-tailed test, rank-order correlation). Apparently there was a tendency for subjects in the front rows to react more in terms of single letters than in terms of the entire word, and this is consistent with the assumption that only parts of the test word were effectively registering.<sup>7</sup>

<sup>6</sup> We are indebted to Leo Hurvich for pointing out this possibility.

<sup>7</sup> Closeness to the screen also seems to restrict the recall of meaning in the absence of partial cues. Recall of BREAD, SMELL, COTTAGE, and BRICK minus flag, gift, mirror, and trunk is correlated .45 with row ( $p < .05$ , two-tailed test). Flag et al. have no struc-

Distance from the screen is, surprisingly, correlated only .15 ( $\rho$ ) with amount of discrimination. Thus we are justified in saying that a decrease in distance did not automatically lead to awareness of the stimulus, and that the factors which contribute to broad effects in the back rows are probably independent of the factors which contribute to good discrimination. The "row effect" probably depends on such general factors as size, intensity and visual angle; the "discrimination effect" on individual factors such as visual acuity, readiness to guess, distribution of attention, etc. Obviously these two sets of factors can interact in a wide variety of ways.

### DISCUSSION

The results show that the meaning of a stimulus can be registered without awareness and can significantly influence subsequent recall. The pattern of recall shown by the Subliminal subjects makes it more parsimonious to conclude that the word CHEESE was registered intact and undistorted, and does not support the assumption that some similar structure was registered which bore only an approximate resemblance to CHEESE. A similar, but not identical, structure would not explain why meaningfully related associates were recalled.

The second conclusion to be drawn from our findings is that reported awareness is not a reliable indicator of the amount of information being registered. Subjects who were unable to discriminate the presence or absence of the test word nonetheless received a substantial amount of meaning information, and seemed relatively uninfluenced by partial cues. Those subjects who could discriminate the stimulus, however, were sensitive to meaning cues, and were not sensitive to meaning. If anything, reported awareness is an indicator that more structurally determined effects are to be expected and the absence of awareness, an indicator that more meaning determined effects are to be expected, as shown by the correlation between degree of

awareness and preference for structure over meaning in the recall.

The results also fail to support the assumption that the response to reduced stimulus input is a crude and pale copy of the response to more intense input. The responses of subjects in the back row of the Subliminal group, receiving the stimulus at substantially reduced intensity, were no less dependent on meaning than were the responses of subjects in the front row. And when information was substantially increased by lengthening exposure speed, the accompanying responses did not include a greater number of CHEESE associates. Subjects in the Supraliminal group received the stimulus about 300 times longer than subjects in the Subliminal group, but the response of the former group was no more complete, as would be reflected in a greater recall of CHEESE associates.

These facts lead us to conclude that the intensity argument is insufficient to account for the results, and take us back to our alternative set of four assumptions which more nearly meet the case. Our findings support Assumption 1 that registration takes place without awareness; Assumption 3 that a subliminal stimulus does not have a weaker effect than a supraliminal stimulus; and Assumption 4 that reduction of awareness changes the patterning of cognitive activity, with different associative processes operating in the two conditions. Assumption 2 no longer holds, and should be rephrased to say that awareness is inversely related to amount of effect.

Before we consider the apparent paradox that an unseen stimulus may have a stronger effect on recall than a clearly visible stimulus, we might consider the pattern of the subliminal effect. It is clear that a reduction of stimulus information does not necessarily break down the structure of what is registered and leave nothing but partial cues to influence the response. Meaning can still be preserved at faint exposure levels, and it fans out along a wide range of predetermined pathways. The pattern of nonconscious generalization does not seem to be influenced by the subject's conscious associations (his commonality) and—rather surprisingly—seems to be inversely related to the Minnesota norms.

tural similarity to the test word. Apparently, the closer the subject is to the screen, the fewer letters of the test word impinge on his fovea and the less meaning is conveyed, thus restricting recall.



When Thorndike-Lorge frequency is held constant, the activation of rare associates seems stronger than of the more popular associates. For some reason the subliminal effect on the popular associates is inhibited and, perhaps as a consequence, it fans out to influence other, more distant associates.

It is also worth mentioning that the subliminal effect, although relatively indirect, is not chaotic. Clearly the test word did not activate a host of unrelated meanings, for if it had there would have been no measurable effect. Activation is directed primarily to words related to the stimulus, and conscious attention is not necessary to keep simple thought processes under control. Preconscious thinking is not only more than a theoretical possibility, but seems to follow the laws of the secondary process, albeit in an inefficient way, for the fanning-out effect, by its very nature, would certainly interfere with the sustained, secondary process thought necessary for solving a particular problem.

This fanning-out effect seems to be greatest in the last two rows of subjects (see Figure 3) where the stimulus falls almost entirely within the acutely sensitive part of the retina. Since all parts of the word must be registered to convey its meaning, the transmission of meaning is more likely when there is complete foveal impingement. As the distance decreases, the number of letters registered in the fovea also decreases and meaning is correspondingly impoverished, probably because the letters which are registered foveally play a disproportionate part in influencing the response, and may bend it away from the category of CHEESE and CHEESE associates and toward words which have only a structural similarity to the test word. In short, the "row effect" may be due to the fact that the fovea is stimulated by the entire word—which conveys meaning—in the back rows, and by single letters—which produce other response alternatives—in the front rows. The correlation between row and tendency to recall meaning associates seems to support this assumption.

So much for distance; what about awareness? Subjects who discriminated better than chance were more sensitive to partial cues than to meaning, a fact which supports the

assumption that single letters somehow inhibit the activation of CHEESE associates. Partial information, since it can be elaborated in a variety of ways, may lead to hypotheses which have little to do with CHEESE and its associates and hence will not improve recall. Fragments of a word seem to force recall of structurally similar words over CHEESE associates with fewer structural similarities. As awareness increased, partial information also increased and tended to restrict the effect on recall.

Thus it would seem that both awareness and wide visual angle may favor registration of single letters over the complete word, and thus lead to a variety of individual hypotheses based on these letters. One might posit three different stages of registration depending on intensity and size of the stimulus. In Stage 1, the stimulus is well below threshold, excites a wide band of association pathways, and produces a broad activation of many conceptual associates. In Stage 2, bits of partial information play a disproportionately greater role and generate hypotheses based on single letters which organize recall. Since these hypotheses can assume a variety of forms, many of which are quite unrelated to the stimulus, they may easily interfere with the expected effect. In Stage 3, the stimulus is supraliminal, clearly within awareness and organizes the response in terms of a particular area of meaning, as shown by the fact that subjects in the Supraliminal condition reported more food associates than subjects in the Subliminal condition. Note that the effect is restricted in Stage 3 for quite different reasons than in Stage 2.

The part played by partial information in Stage 2 may account for the paradox noted earlier in the paper—the fact that the effect of a verbal stimulus appears to diminish as it reaches threshold. In the study by Smith et al. (1959), a subject might elaborate the stimulus ANGRY into angel as it approached threshold, and describe the face in positive terms. Any increase in idiosyncratic responses such as this would tend to dilute the expected effect, and it is important to note that individual parameters are most likely to organize the response as soon as bits of information begin to emerge into awareness. Partial information, for this reason, may be more mislead-

ing than no information at all, and reports based on partial information may be erroneous not simply because less information is presented, but because this information is organized around an idiosyncratic premise that may be only distantly related to the stimulus.

This, then, is the restricting effect of awareness and this is why a subliminal stimulus, well below the threshold of partial recognition, has a more widespread and uniform effect. Randomly discriminating subjects in the Subliminal group remembered more CHEESE associates; Spence and Bressler (in press) found that the effects produced by a stimulus far below threshold conformed more exactly to the Minnesota norms than the effects produced by a stimulus closer to threshold. In both cases, as the stimulus moved closer to threshold it gave rise to individual response tendencies (probably mediated by partial cues) and the accompanying response was stamped by idiosyncratic earmarks.

In view of these findings, we must take issue with Eriksen's (1960) recent statement that "there is no convincing evidence that the human organism can discriminate or differentially respond to external stimuli that are at an intensity level too low to elicit a discriminated verbal report" (p. 298) and say that a stimulus which is not verbally discriminated is nevertheless capable of remarkably accurate and far-reaching effects which involve meaning and stand in contrast to the structural effects mediated by partial cues.

#### SUMMARY

The purpose of this experiment was to study how level of awareness influenced the effect of a verbal stimulus on recall. Sixty-five undergraduate students were presented a list of words to recall, half of which were associates to the test stimulus CHEESE, and half were not. Before the list was read, the stimulus was presented subliminally to one group, supraliminally to a second, and never exposed to the third, Blank group. Whereas both Blank and Supraliminal groups showed no preferential recall of CHEESE associated words, the Subliminal group showed a significantly greater recall. Degree of effect was inversely related to awareness of the stimulus, and to closeness of the stimulus. Further

analysis of the role of level of awareness led to the postulation of three stages of registration: Subliminal with no partial cues; Nearliminal with partial cues available; and Supraliminal. More meaning associates were recalled in Stage 1; more structurally related words in Stage 2. The features characterizing each stage were discussed in detail and led to the conclusion that an explanation based on partial cues may explain Stage 2, but not Stage 1 for which additional assumptions are necessary.

#### REFERENCES

- BRESSLER, J. Illusion in the case of subliminal visual stimulation. *J. gen. Psychol.*, 1931, 5, 244-250.
- DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, 5, 305-312.
- EAGLE, M. The effects of subliminal stimuli of aggressive content upon conscious cognition. *J. Pers.*, 1959, 27, 578-600.
- ERIKSEN, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychol. Rev.*, 1960, 67, 279-300.
- FISHER, C. A study of the preliminary stages of the construction of dreams and images. *J. Amer. Psychoanal. Ass.*, 1957, 5, 5-60.
- FOX, MURIEL. Differential effects of subliminal and supraliminal stimulation. Unpublished doctoral dissertation, New York University, 1960.
- FREUD, S. *The interpretation of dreams*. (Originally published 1900) Vols. 4 & 5. (Trans. by J. Strachey) London: Hogarth, 1953.
- GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, 55, 373-411.
- JENKINS, J. J., & RUSSELL, W. A. Miscellaneous studies in word association. Technical Report No. 24, 1958, University of Minnesota, Contract N8 onr 66216.
- KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNAH. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, 57, 255-266.
- LACEY, J. I., & SMITH, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, 120, 1045-1052.
- LAPKIN, B., & LIPPMANN, P. Is there a need for two concepts: Perception and registration? Unpublished manuscript, New York University, Research Center for Mental Health, 1959.
- LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, 58, 113-122.
- PAUL, I. H., & FISHER, C. Subliminal visual stimulation: A study of its influence on subsequent im-

- ages and dreams. *J. nerv. ment. Dis.*, 1959, 129, 315-340.
- PINE, F. Incidental versus focal presentation of drive related stimuli. *J. abnorm. soc. Psychol.*, 1961, 62, 482-490.
- RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr 66216.
- SMITH, G. J. W., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, 59, 167-176.
- SPENCE, D. P. The multiple effects of subliminal stimuli. *J. Pers.*, 1961, 29, 40-53.
- SPENCE, D. P., & BRESSLER, JOANNA. Subliminal activation of conceptual associates: A study of rational preconscious thinking. *J. Pers.*, in press.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Bureau of Publication, 1944.
- WEINER, M., & SCHILLER, P. H. Subliminal perception or perception of partial cues. *J. abnorm. soc. Psychol.*, 1960, 61, 124-137.

(Received March 1, 1961)



## SOME EFFECTS OF REJECTION UPON ATTRACTION TO A GROUP<sup>1</sup>

J. DIEDRICK SNOEK

Research Center for Group Dynamics, University of Michigan

The process by which individuals move in and out of groups can be thought of as a mutual selection during which the individual evaluates the attractiveness of membership in a particular group and the group judges his acceptability as a member. Among the factors of interest to social psychologists in this process are the motivational basis for the individual's initial interest in becoming a member, the criteria by which the groups accept or reject him, and finally the basis for his continued participation in the group. While previous studies have paid considerable attention to the bases for an individual's attraction to membership, given his acceptance into a group, few studies have concerned themselves with changes in the nature of an individual's desire for membership in a group as a result of his acceptance or rejection. The study to be reported below concerns some conditions under which an individual will continue to be attracted to membership in a group which has rejected him.

The experience of rejection has two important psychological aspects. First, rejection deprives the individual of the opportunity to gratify whatever needs led him to desire membership in the group. Secondly, rejection represents an evaluation by the group of the rejected person. To the extent that the criteria used by the group in making its evaluation are relevant to the individual's self-evaluation, rejection may be more or less damaging to self-esteem. In this paper we shall speak of *invidious rejection* whenever exclusion from a group can be taken to mean that the individual is not worthy of membership in it. We shall call the rejection *non-invidious* when it is based on other reasons

that do not reflect on the individual's self-esteem; when an individual is excluded because all membership positions in a group are filled we have an example of noninvidious rejection.

Just as it is possible to distinguish various degrees of acceptance in a group, various degrees of rejection are possible. Several experiments have tried to manipulate this variable, although conceptually it has not always been clear whether "degree of rejection" referred to the amount of frustration experienced by the rejected person or to the negativity of the evaluation he received. In this paper the concept of degree of rejection will refer to the strength of the group members' determination to resist the individual's attempts to participate as a member in the group, after Jackson's (1959) usage.

The concept of attraction to a group often is taken to refer to the strength of an individual's desire to participate in a group without regard to the particular needs that may motivate him. In this paper attraction will be considered a joint function of the valence of membership in the group and the expectation that membership can be attained, which can be expressed as the individual's subjective probability of acceptance. That is to say that in a choice situation the individual's tendency to choose membership in a group will be a function of both the apparent attractiveness of the group and his expectation that he can join it (cf. Feather, 1959). For convenience it is assumed that the function relating valence and subjective probability of acceptance is a simple multiplicative one.

Given this formulation of attraction, two hypotheses can be formulated regarding the expected effects of rejection upon attraction to a group. First, whatever the reasons for rejection, the individual will revise his estimate of the probability of his acceptance

<sup>1</sup> This paper is based upon a doctoral dissertation submitted to the Graduate School of the University of Michigan in 1959. The author wishes to acknowledge his gratitude to the Social Science Research Council for the fellowship which permitted him to conduct this experiment.

downwards to an extent that is proportional to the degree of rejection. The result of such a change would be a decrease in his overall attraction to the group. The second effect to be expected is a change in the valence of membership depending on whether or not the rejection was invidious. Jackson and Snoek (1959) found that the reactions to rejection of very strongly rejected subjects, as contrasted with those who only experienced mild rejection, appeared to be motivated by a need to reassure themselves about their acceptability to others (need for social reassurance). The strongly rejected subjects maintained their attraction to the group, while subjects in the mild rejection conditions appeared to have decreased theirs. In the present study it is expected that the need for social reassurance will be strongly aroused in invidiously rejected subjects who in turn will value reacceptance into the group as a means of reassuring themselves. Accordingly, the valence of membership is expected to increase when the reasons for rejection are invidious; this change in valence is assumed to be proportional to the degree of rejection.

In summary, invidiously rejected subjects should experience two opposing tendencies to change their attraction to the group, one to lower their attraction due to a decrease in their estimate of the probability of acceptance and a second tendency to increase their attraction due to their need for social reassurance. By contrast noninvidiously rejected subjects will be subject only to the first tendency to lower their attraction to the group. From these assumptions it is possible to predict the following differences in relative changes in attraction to a group after the two kinds of rejection:

*Hypothesis 1.* For any given degree of rejection, noninvidious rejection will be followed by a greater reduction in attraction to the group than invidious rejection.

*Hypothesis 2.* Differences between invidious and noninvidious rejection conditions in loss of attraction to the group will be greater after strong than after mild rejection.

## METHOD

*Subjects.* Experimental groups were organized by asking male undergraduates at the University of

Michigan to join a series of bridge clubs, organized for purposes of experimental observation as well as for their enjoyment. Fifty seven subjects participated in the experiment. Subjects were both recruited and scheduled for meetings by the author, identified as a graduate assistant at the Research Center for Group Dynamics.

*Procedure.* The usual group consisted of two subjects and two hired experimental assistants, one male and one female, who pretended to be subjects. At the time of scheduling the experimenter told each subject he would be meeting with an already established group which lacked a fourth member; in this way subjects were put in the position of being candidates for the position of "fourth member" in an already existing group. Since subjects generally arrived at the meeting at slightly different times, it was possible to introduce each of them separately to the two assistants and then to each other, so that each subject remained unaware that the other was also a newcomer to the group.

After the introductions the members of each group were told to proceed as they wished, while the experimenter took the passive role of observer. Most groups immediately started to play bridge, using a set of predealt hands arranged so as to keep each subject's and group's bridge experiences comparable; groups which spent more time getting acquainted were prodded into playing bridge after a short time by one of the experimental assistants. After approximately 1 hour, the experimenter requested an interruption in the meeting in order to administer a questionnaire. For this purpose subjects were separated into cubicles, a procedure which allowed the experimenter to administer the subsequent rejection induction to each of the two subjects in private. After the rejection manipulation subjects rejoined the group for another 45 minutes of bridge playing, at the end of which they were again asked to fill out a questionnaire. After the second questionnaire, subjects were reassembled, told of the experimental manipulations and given a chance to share their reactions with the other participants.

Four experimental conditions were created: Strong Invidious, Mild Invidious, Strong Noninvidious, and Mild Noninvidious rejection. In all four conditions rejection was induced by the experimenter under the pretense of feedback of spurious questionnaire ratings subjects had made of each other. The following instructions were used:

Because you are in the unusual position of being a newcomer to this group and because the evaluations you received look somewhat different from those of other members in this group, I thought I would discuss these ratings with you now rather than wait till the end of the session, as I usually do. Do you remember that all of you were asked to write on the first questionnaire whether you preferred to continue playing with each other in future meetings? As you can see [showing spurious evaluations] most of the people



in this group agree that they want to continue playing with each other.

[*Strong rejection treatment*] But as you can see, their rating of you is quite a bit lower than that of the others. In other words, most of the members of this group seem to prefer that you do not try to join this group.

[*Mild rejection treatment*] But as you can see, their rating of you is not quite as high as that of the others. In other words, the other members of this group seem to be neither in favor of nor opposed to your joining the group.

[*Invidious reason for rejection*] As the other ratings show, you are apparently considered not quite as good a bridge player as they would have liked; you are also not as well-liked as some others in this group. Apparently, they seem to think that you would fit rather badly into this group.

[*Noninvidious reason for rejection*] As the other ratings show, you are considered just as good a bridge player and just as likeable a person as anyone else in the group. What seems to be the case, however, is that they would prefer to have a second girl playing with them in the future, rather than having three men and only one girl as would be the case if you became the fourth member.

After these instructions subjects were enjoined to finish the questionnaire on which they were working and then to rejoin the group. Meanwhile the experimenter would repeat the rejection manipulation with the second subject in another cubicle. On a few occasions the second subject would already be out of his cubicle and no opportunity remained to give the rejection instructions without arousing suspicion; five subjects therefore received no experimental inductions. Four subjects questioned the rejection induction, stating that they thought it was an experimental manipulation; these subjects were eliminated from the analysis.

*Questionnaire measures.* Attraction to the group was measured with a set of three questions, taken from Libo (1953). Attraction scores based on summing over the three items showed a correlation of .92 (biserial  $r$ ) with Libo's "locomotion criterion" of choosing to stay in or to leave the group (computed by the author on Libo's data). The three questions were: "Do you want to remain a member of this group?" (Desire to Stay), "How often would you like to come to meetings of this group?" (Frequency of Meeting), "If this group broke up for a considerable length of time and some people were trying to get it started again, would you want to rejoin?" (Desire to Rejoin).

A measure of Subjective Probability of Acceptance was obtained from answers to the question: "How likely do you think it is that the other members of this group want you to remain a member?" The question had five alternative answers, ranging from "very likely" to "very unlikely." Like the attraction

questions, this question was repeated after the rejection manipulation and the second period of interaction with the group.

In addition, the final questionnaire included an item on Attraction to Membership in Another Group ("Instead of continuing to meet with this group, would you like to have the opportunity to meet with another group like this?") and one on Desire for Contact with individual members of the group outside of the experimental sessions, based upon a seven-item scale constructed by Back (1951).

*Observation measures.* Verbal interaction was recorded throughout the two periods of interaction, before and after the rejection manipulation. A record was kept of (a) the frequency with which each subject spoke, (b) the content of his remarks in terms of a very simple classification schema, and (c) the passage of time. The unit of speech coded was a *remark*, defined as a series of phrases or sentences, uttered without interruption, about a single topic. If the topic was changed once or more often during a single utterance, the subject was recorded as having initiated two remarks. Content was coded as either (a) *task demanded*, defined as remarks that were unavoidable in the context of the bridge game, such as bids or responses; (b) *task related*, defined as spontaneous remarks related to but not required by the bridge game; (c) *personal*, defined as remarks about the player himself or about one of the other players' background or attributes; or (d) *other*, a residual category. Pretesting showed that the great majority of remarks during the bridge game could be coded reliably into the first three categories. The reliability of the coding was estimated by correlating the rates of participation for each content category and participant obtained by two observers over nine observation periods of approximately 10-minutes duration. The obtained correlation coefficient was .86.

*Role of experimental assistants.* The two assistants played no role in the rejection manipulation, but were used only because it was impossible for the experimenter to deal with more than two subjects at a time. Their familiarity with the room and procedures reinforced the subjects' impression of joining an already "established" group. Because subjects for each group were selected on the basis of self-ratings of their bridge competence, the assistants were instructed to simulate a level of competence comparable to that of the subjects. The competence ratings received by the assistants showed they were indeed always considered to be roughly equal in competence to the subjects. Similarly, the assistants were instructed to initiate as little interaction as possible on their own, but instead to imitate as closely as possible both in content and volume the patterns of interaction their partners (the subjects) seemed to prefer, so that the interaction rates might reflect as closely as possible the subjects' preferences. Analysis showed that the assistants were reasonably successful in this attempt; their participation rates over all experimental sessions correlated .40 and .66 with their respective partners' scores.



TABLE 1

NUMBER OF SUBJECTS CHANGING PROBABILITY OF ACCEPTANCE ESTIMATES AFTER REJECTION

Probability of acceptance	Rejection conditions			
	Mild invidious	Mild noninvidious	Strong invidious	Strong noninvidious
No change	8	7	4	4
Decreased	4	5	8	8

Note.—Mild vs. Strong:  $\chi^2 = 4.16$ ,  $p < .05$ ; Invidious vs. Noninvidious:  $ns$ .

## RESULTS

Due to the loss of 4 subjects who discovered the deception and 5 subjects who never received the rejection treatment, the analysis concerns 48 subjects who were randomly assigned to four experimental conditions, 12 to each condition.

The effectiveness of the strength of rejection instructions could be checked by comparing the relative decrease in subjective probability of acceptance for the mild and strong rejection conditions. Table 1 shows that the strong rejection instructions led to a reliable increase in the number of subjects who decreased their estimate of the likelihood of being accepted by the group. Reason for rejection, as expected, had no discernible effect upon changes in subjective probability of acceptance.

Prior to the experiment subjects returned a mailed questionnaire, consisting of 20 seven-point scales, on which they had described the kind of group they would prefer to join. After the first interaction period subjects used the same instrument to describe the actual groups they had participated in. Assuming that congruence between preferred and actual descriptions indicates a maximally attractive group, a measure of the groups' potential to satisfy the subjects' interests was computed from discrepancies on the three scales which each subject had called "most important" in making the group attractive to him (Satisfaction Potential).

In order to test the validity of the theoretical assumption that attraction would be a function both of the valence of membership and subjective probability of acceptance, correlations of each of the latter two variables

with the prerejection Desire to Stay scores were obtained. Subjective Probability of Acceptance correlated .23 (biserial  $r$ ,  $ns$ ) and Satisfaction Potential .55 ( $p < .01$ ) with Attraction to the Group. Scores obtained by multiplying the Subjective Probability and Satisfaction Potential variables correlated .59 ( $p < .01$ ) with Desire to Stay. While combining the two variables thus seemed to add very little in predictive power to the Satisfaction Potential measure taken singly, inspection of Table 2 indicates that the addition of the Subjective Probability of Acceptance variable made a considerable difference when Satisfaction Potential scores were high. The validity of our assumption about the determinants of Attraction to the Group therefore appeared partially confirmed.

The before and after means for each of the three attraction questions in the four experimental conditions are listed in Table 3, together with the mean changes in attraction after rejection. The average level of attraction to the group was moderately high before rejection and did not differ significantly between experimental conditions. Two of the attraction measures, Desire to Stay and Frequency of Meeting, show a similar pattern of changes after rejection. On both measures, Strong Noninvidious rejection led to a decrease in attraction to the group, while neither of the Invidious rejection conditions produced any change.

The analysis of variance of these results, summarized in Table 4, indicates that the effect of reasons for rejection was significant at the 5% and 10% levels of confidence, respectively, for the Desire to Stay and Fre-

TABLE 2  
MEAN DESIRE TO STAY BY SATISFACTION POTENTIAL  
AND SUBJECTIVE PROBABILITY OF ACCEPTANCE  
( $N = 57$ )

Satisfaction potential	Subjective probability acceptance	Mean desire to stay	$N$	$SD$	$t$
High	High	8.71	17	1.12	8.20****
	Low	7.37	19	1.84	
Low	High	5.82	11	3.56	0.37
	Low	5.30	10	2.43	

\*\*\*\*  $p < .001$ .

TABLE 3

MEAN SCORES ON VARIABLES RELATED TO ATTRACTION BEFORE AND AFTER REJECTION

Variable	Rejection conditions			
	Mild non-invidious	Strong non-invidious	Mild invidious	Strong invidious
Desire to Stay				
Before	7.6	7.1	6.5	7.3
After	6.4	5.0	6.4	7.4
Change	-1.2	-2.1	-0.1	0.1
Frequency of Meeting				
Before	3.7	3.8	3.8	3.8
After	3.6	3.1	4.1	3.8
Change	-0.1	-0.7	0.3	0.0
Desire to Rejoin				
Before	6.7	6.9	6.3	8.0
After	6.7	5.7	6.4	6.8
Change	0.0	-1.2	0.1	-1.2
Desire for Contact				
After	4.2	3.8	4.6	3.3
Attraction to a New Group				
After	7.3	9.6	6.0	7.9

quency of Meeting measures.<sup>2</sup> The third measure of attraction, Desire to Rejoin, shows a different pattern of results, however. Reasons for rejection produced no effect at all on this measure, but degrees of rejection apparently produced differences sufficient to reach the 10% level of confidence. As Table 3 shows, there was no loss of attraction after mild rejection on this measure, but strong rejection led to some decrease in desire to rejoin the group. In summary, Hypothesis 1 was supported by data on changes in attraction to the group on two of our measures, but not on a third.

The differences in the results for the three measures unfortunately cannot be accounted for by lack of correlation between them. Desire to Rejoin correlated .74 ( $p < .01$ ) with Desire to Stay and .65 ( $p < .01$ ) with Frequency of Meeting; the latter two measures correlated .83 ( $p < .01$ ) with each other ( $N = 57$ ). Possibly the differences that appeared after the rejection treatment can be

accounted for by the time reference of the questions. Desire to Stay and Frequency of Meeting refer to present intentions to continue to participate in the group, but Desire to Rejoin refers to future participation, which might be much less relevant to the subjects' present need for social reassurance.

According to Hypothesis 2 we would expect an interaction effect between reasons for and degrees of rejection. Although on the Desire to Stay and the Frequency of Meeting measures of attraction the differences in attraction loss are indeed greater for the two strong rejection conditions than for the mild rejection treatments, the predicted interaction effect does not occur. It appears therefore that Hypothesis 2 must be rejected on the basis of our data.

Two other results displayed in Table 3 remain to be discussed. Desire for Contact outside the experimental sessions was thought to reflect the subjects' liking for each other as persons rather than as members of the same bridge club. Although the means for the mild rejection treatments were slightly

TABLE 4  
ANALYSIS OF VARIANCE SUMMARIES FOR VARIABLES RELATED TO ATTRACTION TO THE GROUP

Variable	Source	MS	df	F
Change in Desire to Stay	Degrees	1.69	1	6.05**
	Reasons	31.69	1	
	D × R	3.51	1	
	Error <sup>a</sup>	5.24	44	
Change in Frequency of Meeting	Degrees	2.52	1	3.29*
	Reasons	3.52	1	
	D × R	0.19	1	
	Error <sup>a</sup>	1.07	44	
Change in Desire to Rejoin <sup>b</sup>	Degrees	0.66	1	3.88*
	Reasons	0.01	1	
	D × R	0.00	1	
	Error <sup>a</sup>	0.17	44	
Desire for Contact	Degrees	80.20	1	2.99
	Reasons	00.40	1	
	D × R	26.70	1	
	Error <sup>a</sup>	26.82	44	
Attraction to a New Group	Degrees	52.10	1	7.54***
	Reasons	27.00	1	
	D × R	00.40	1	
	Error <sup>a</sup>	6.91	44	

<sup>a</sup> Error term based on pooled estimate.

<sup>b</sup> Scores adjusted by square root transformation to correct for heterogeneity of variance.

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

<sup>2</sup> In order to rule out the possible effect of variations in attraction to the group before rejection, the data were also analyzed by the analysis of covariance method, controlling on prerejection attraction. The results of this analysis left the  $F$  values for the attraction measures essentially unchanged from those reported in Table 4.

TABLE 5  
MEAN CHANGE IN PARTICIPATION RATES BY CONTENT  
AND EXPERIMENTAL CONDITIONS\*

Content area	Rejection conditions			
	Mild noninvidious	Strong noninvidious	Mild invidious	Strong invidious
Task demanded	0.2	-0.3	-0.8	-1.3
Task related	-0.1	-0.1	-0.7	-0.8
Personal	-1.5	-1.8	-0.1	0.3
Total change	-1.4	-2.2	-1.6	-1.8

\* Negative numbers indicate decreases in rates of participation.

higher than those for the strong rejection conditions, the difference between them is not statistically significant. The rejection treatments did not create any differences in desire to affiliate with the members of the bridge clubs, but this result can probably not be interpreted to mean that the rejection treatments produced no hostility towards members of the rejecting group. The experimenter was told of hostile feelings towards others in the group by a number of subjects during the post-experimental discussions.

Subjects were also asked to indicate their preference for joining another bridge club than the present experimental group. As the results for Attraction to Another Group show, strong rejection appeared to increase the subjects' willingness to join another group more than mild rejection in both invidious and noninvidious conditions ( $p < .01$ , Table 4). Reasons for rejection may also have had some effect, however; invidious rejection appears to depress the subjects' willingness to join another group slightly ( $p < .10$ , Table 4). It is of interest to note that the mean for attraction to another group in the Strong Noninvidious condition is greater than the mean for attraction to the experimental group before rejection. Apparently strong rejection tends to boost the subjects' interest in joining a bridge club to a higher absolute level.

Rates of verbal participation were calculated for the interaction period before and after rejection. Table 5 presents mean changes in rates of participation, broken down by content area. In general, participation decreases by about the same amount in all experimental conditions. The content area in which the

decrease occurs, however, is quite different for the Invidious and Noninvidious conditions. Noninvidiously rejected subjects decrease their participation by reducing the number of *personal* remarks they make after rejection; their rates of participation in the *task demanded* and *task related* content areas remain virtually unchanged. The effect of reason for rejection on changes in personal content rates is significant at the 1% level of confidence (Table 6). Invidiously rejected subjects do not reduce personal remarks, as Table 5 shows, but do reduce the number of their remarks in the task demanded and task related categories; this latter effect falls short of statistical significance. If we interpret rate and content of participation as indicators of the subjects' motivation to participate in the group, it appears that noninvidious rejection leads to a withdrawal of personal interest in the group, although interest in the task remains unaffected. This kind of withdrawal fails to appear after invidious rejection, lending support to Hypothesis 1 about differences in tendencies to withdraw from the rejecting group between invidious and noninvidious rejection conditions.

DISCUSSION

It should be noted that the manipulation of degree of rejection by itself failed to reduce attraction to the group on two out of three measures, unlike the results of earlier experiments which showed a direct relation between lack of acceptance and resultant attraction to the group (Dittes, 1959; Dittes & Kelley, 1956; Kelley & Shapiro, 1954). The explanation may be found in differences between the earlier experiments and this one in the degree of commitment to a choice that

TABLE 6  
ANALYSIS OF VARIANCE SUMMARY FOR CHANGES IN  
PERSONAL CONTENT OF COMMUNICATIONS

Source	MS	df	F
Degrees	0.02	1	9.85***
Reasons	38.52	1	
D X R	1.69	1	
Error*	3.88	44	

\* Error term based on pooled estimate.  
\*\*\*  $p < .01$ .



was implied in the attraction measurement procedures. Feather (1959) has pointed out the need to differentiate between measures of "attainment attractiveness, defined in terms of a wishful choice in a situation where there is no implied commitment," and "choice potential, defined by a choice in a situation where there is implied commitment." The procedures used in the present study definitely led subjects to believe they had to choose whether or not to attend future meetings of the group. As was seen above, the measure that is least like a direct choice because it refers to a hypothetical future event (Desire to Rejoin), produced results that at once differed from the other two attraction measures and were most like the results reported in earlier studies. In short, the results of this study conformed to the hypothesis about the effects of different reasons for rejection where measures of *choice* of participation in the group were involved (i.e., Desire to Stay and Frequency of Meeting). Desire to Rejoin the group in the future, which is more likely to measure *preference* for membership in the group without commitment to an actual choice, was not influenced by reason for rejection but only by the degree of rejection. If one takes the lack of Desire to Rejoin on the part of both strongly rejected groups to mean that membership in the group was actually unattractive to both of them, it can be concluded that the reason invidiously rejected subjects continued to choose membership in the rejecting group was probably that nonmembership had become more repulsive to them as a result of their invidious rejection.

It is likely that invidiously rejected subjects will not continue to choose membership in the rejecting group under all circumstances. In terms of the theoretical schema presented above, it is to be expected that over time (a) subjective probability of acceptance would decrease until the resultant attraction to the group approaches zero, and that (b) as the individual continues to experience rejection the valence of membership will in fact be lowered. In this connection it should be noted that the differences obtained in this study between invidious and noninvidious rejection cannot be explained in terms of differential expectations about being reaccepted into the

group. The data do not show, for example, that noninvidious rejection led to a greater decrease in Subjective Probability of Acceptance than invidious rejection (Table 1). The evidence does support the view that reason for rejection is an important factor in determining the subjects' reactions to rejection. In particular, the more invidious the grounds for rejection, the more likely it is that new needs come into play that help determine the subjects' behavior. Under circumstances such as were employed in this study such needs may cause subjects to prefer efforts to join the rejecting group over leaving.

#### SUMMARY

A distinction was made between invidious rejection, defined as rejection for reasons that are likely to reflect upon the subjects' self-esteem, and noninvidious rejection. On the basis of the hypothesis that invidious rejection would arouse a need for social reassurance, it was predicted that (a) for any given degree of rejection, loss of attraction to the group will be smaller when the reason for rejection is invidious than when it is noninvidious, and (b) that the differences in loss of attraction between invidious and noninvidious conditions will be greater after strong than after mild rejection.

The predictions were tested in a  $2 \times 2$  factorial experiment, employing two degrees of rejection and two different reasons for rejection, one invidious and the other noninvidious. Both attraction to the group and subjective probability of acceptance were measured before and after an experimentally induced rejection experience. The groups were experimentally created bridge clubs of four members, two of whom were confederates of the experimenter. There were 48 experimental subjects, all college students. The results with respect to attraction to the group generally supported Hypothesis 1, but not Hypothesis 2. In addition it was found that in the noninvidious condition subjects engaged in significantly less conversation of a personal nature after rejection. A secondary finding of interest was that willingness to join another group was greater after strong than after mild rejection, exceeding even the original level of attraction to the experimental group.

## REFERENCES

- BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, 46, 9-23.
- DITTES, J. E. Attractiveness of group as function of self-esteem and acceptance by group. *J. abnorm. soc. Psychol.*, 1959, 59, 77-82.
- DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance on conformity to group norms. *J. soc. Psychol.*, 1956, 53, 100-107.
- FEATHER, N. T. Subjective probability and decision under uncertainty. *Psychol. Rev.*, 1959, 66, 150-164.
- JACKSON, J. M. A space for conceptualizing person-group relationships. *Hum. Relat.*, 1959, 12, 3-16.
- JACKSON, J. M., & SNOEK, J. D. Effect of invidious exclusion from a group on feelings toward self, others and on tendencies to conform. Paper presented at the sixty-seventh Annual Meetings of the American Psychological Association, Cincinnati, Ohio, 1959.
- KELLEY, H. H., & SHAPIRO, M. An experiment on conformity to group norms where conformity is detrimental to group achievement. *Amer. sociol. Rev.*, 1954, 19, 667-677.
- LIBO, L. M. *Measuring group cohesiveness*. Ann Arbor: Research Center for Group Dynamics, 1953.

(Received January 20, 1961)

## SIMILARITY OF PERFORMANCE AS INFLUENCED BY INTERACTION, SELF-ESTEEM, AND BIRTH ORDER<sup>1</sup>

EZRA STOTLAND AND NICKOLAS B. COTTRELL

*University of Washington*

This study concerns the relationship between the amount of interaction between two persons in a group and the acquisition by one of the persons of an attribute of the other. It is hypothesized that the greater the interaction between the two, the more will one acquire an attribute of the other. In this fashion the effects of interaction per se can be studied, that is, the effects of interaction in which the information communicated by the interaction has no relevance to the outcome or dependent measure studied. This study also investigates the differential tendencies of group members of high and low self-esteem to acquire attributes in this manner.

The work of Michotte (1946) provides the theoretical source of this hypothesis. His studies of the perception of causal relationships between physical objects suggest that when one object is perceived as causing a change in a second object, the change in the second is perceived to belong to the first. In other words, a new unit or new entity is perceived, consisting of the first, causing object and of the change in the second object. This proposition has been extended theoretically to the area of social psychology by Heider (1958). He postulates that a person and the perceived consequence of his behavior will be seen as a unit.

The theory can be extended to the area of the perception of the interaction between two (or more) persons. By definition, when two people are interacting, each act by one of them is followed by, or causes, some event (e.g., change in behavior) in the other. The causing person and the event in the other person would then constitute a perceptual unit. For example, a drill sergeant barks a command at his squad of soldiers and they respond immediately. The sergeant and the

squad would be perceived by an observing officer, as well as by the soldiers and the sergeant, as constituting a single unit.

One additional assumption is needed to derive our hypothesis. The phenomenon of assimilation found in the study of the perception of physical objects is the tendency to perceive all parts of the same perceptual unit as having the same quality, e.g., degree of brightness. It is reasonable to expect that this assimilative tendency toward perception of the uniformity of parts of a unit would be greater in strong units than in weaker units. Therefore, it would be expected that the stronger the composite unit formed by the two persons interacting, the greater will be the tendency for the two persons to be perceived as having the same attributes. These similar attributes do not have to have anything directly to do with the quality, type, or nature of the interactions upon which the composite unit is based. In our example, the sergeant may be perceived to be an excellent marksman, so that his men come to raise their evaluation of their marksmanship; the composite unit was based on his giving of orders on the drillfield, while the similar attribute, marksmanship, might be perceived only on the rifle range.

This relationship between interaction and perception of similarity of attributes has already been demonstrated for the case in which interaction leads one of the persons to perceive the other as being very similar to himself by attributing his own traits to the other (Stotland, Cottrell, & Laing, 1960). The present study concerns one person's acquiring an attribute he initially perceives in the other. The main hypothesis is that interaction between two persons will result in perceived similarity of their attributes.

The particular attribute (equivalent to marksmanship in the above example) em-

<sup>1</sup> This research was conducted under a grant from the National Institute of Mental Health (M 2433).



ployed to test this hypothesis in the present study is self-evaluation of performance on a task in a laboratory setting. The hypothesis applies to the individual's perception of his ability on the task. However, the effects of acquisition of this attribute may be deep enough to influence an individual's actual level of performance when he is given an opportunity to perform on the same task. A second hypothesis then predicts that when two persons have interacted to a high degree, the level of performance of one will directly influence the level of performance of the other more than when they have interacted to a low degree.

It is also assumed on the basis of a study by Stotland et al. (1960) that the boundaries of the self-conceptions of persons low in self-esteem are more penetrable than of persons high in self-esteem. In the present study, the penetration would be the person's perceiving himself to have the same level of ability on a task as a person with whom he has interacted. Thus, the third hypothesis of the present study is that persons with low self-esteem will behave more consistently with respect to the first two hypotheses than will persons high in self-esteem.

### METHOD

*Subjects.* Students attending introductory and intermediate courses in psychology at the University of Washington during Fall Quarter, 1959, served as subjects. They were recruited at the beginning of the quarter to serve in a program of four experiments of which this research was one. The subjects received payment of \$5.00 for serving in all four of the studies in the series.

The subjects were combined into six-person, same-sexed groups. An attempt was made to minimize prior acquaintance. Hence group members were chosen from different classes and within group similarities in last names and addresses were avoided.

On some occasions one or more of the scheduled subjects failed to appear at the scheduled time. At these times the experimenter summoned a paid assistant and ushered him in as if he were 'he tardy subject. These "stooges" had no special function. They followed instructions like the other subjects. The data gathered from these "stooges" were not considered in the analysis.

*Procedure.* When a group of six subjects arrived at the experimental room they were seated at a round table in the middle of which was a Lazy Susan with six compartments. The table was divided into six sections by fiberboard separators;

thus, each subject could see only the tops of the other subjects' heads. On the table space in front of each subject were: an envelope containing 10 or 12 packs of eight syllables each, a 3"  $\times$  5" pad of pages, a box of paper clips, a pencil, a 3  $\times$  5  $\times$  3 wooden file box, and a record sheet. The subjects were told that they were first to form words from their packs of syllables. Each eight-syllable pack yielded four words. The subjects fastened pairs of fitting syllables together with paper clips. The words formed fell into one of six categories. These categories were animals, birds, cities of the United States, foreign countries, fruits and vegetables, and modes of transportation. The subjects' second job was to collect words of one category. The category assigned to them appeared on their file box. When they formed words in their own category, they put them in their file box. Some of the words they formed belonged in categories other than their own. These words they sent off to the other categories by means of the Lazy Susan. The Lazy Susan was divided into six sections and each section was labeled with one of the category headings. To send a word to another category the subject turned the Lazy Susan until he found the category he wanted to send to, then he put on the word and turned the Lazy Susan back until his own category again confronted him. Subjects were instructed to keep track on their record sheet of the number of words they sent to each other category and the number of words they received from each category. When the subjects sent a word they were instructed to attach a slip of paper with the name of their category written on it.

The final instruction to the subjects was that there was to be no talking during the study.

The only interaction the subjects had with each other was the sending and receiving of words. It is obvious that the distribution of words could be arranged so that varying patterns or amounts of word interchange would result. In this study the materials were arranged so that every person sent eight words to the animals category and received eight words from animals. All categories but animals formed 40 own words in addition, forming a total of 48; the animals category formed none of its own words, but did form 8 for each of the five other types of words, from a total of 40.

Each group worked for 20 minutes on the interaction task. Most subjects, but not all, finished all their word assembling. After this they were told they were to try a new task. First, the experimenter had one of the group members demonstrate the new task. In the Interaction condition the demonstrator or model was the person who had collected the animal words and hence had been the center of the interaction (interchange of names). In the No Interaction condition the demonstrator was a person from the periphery of the prearranged interaction network; no one but the collector of animal words had exchanged words with the demonstrator in this condition. Groups were alternately assigned to these conditions.

The second task consisted of listening to a tape recorded reading of a passage in a foreign language (the language was Hindi) and keeping a mental count of the number of times a certain sound occurred. There were two passages and the person's score was the sum of the errors on each passage.

The demonstrators were first given the tape task; when it was over, the demonstrator wrote down his count. The demonstrator listened to the tape over a set of earphones so that the other subjects could not hear the tape while he was listening to it. In half the cases the demonstrator was told in the presence of the group that he had been very accurate in his count, his count being only 15% away from perfect. This accuracy, the experimenter said, was proof of his paying close attention.

This condition was called the Good Evaluation of Model's Performance. In the other half of the sessions, the demonstrator was told that he had been rather inaccurate in his counting, his count being 85% away from perfect. This inaccuracy, the experimenter said, was probably due to not paying attention. This condition is called the Poor Evaluation of the Model's Performance. Groups were assigned alternately to these two conditions.

The other subjects were then asked to rate how well they expected to do on the task. Then the subjects were given their own sets of earphones and listened to the two passages, recording their mental count at the end of each selection. They then rated their own performances. The subjects then indicated on a questionnaire how well they had known each of the other subjects, their level of interest in the listening task, and attractions to the group qua group, and desire to know each of the other group members. All of these ratings were made on seven-point scales.

The reasons that this task was selected were, first, that it was completely unfamiliar to the subjects and they would have no standards available upon which to base expectations of own performance; second, the model's performance would be judged only by the experimenter; third, the subjects

could all perform the task simultaneously without knowledge of each other's level of performance; fourth, previous research using this task indicates that performance on this task is susceptible to social influence.

The data from subjects who were the center of the interaction network or were demonstrators were discarded from the analysis. At a later session, the study was explained to the subjects; also, their prior training in Hindi was ascertained. Any subjects who had prior experience with Hindi were dropped from the sample. Any subjects in the No Interaction condition who by mistake interacted with the demonstrator were dropped from the sample. Subjects who indicated they had known their demonstrator before serving in the study were also dropped from the sample. The experimental sessions lasted about an hour.

During a previous session of the four in which the subjects participated, they took a modification of the Q sort to measure self-esteem. This measure is described in Stotland et al. (1960).

## RESULTS AND DISCUSSION

*Total sample.* The general hypothesis of this study is that the Interaction condition would lead the subjects to acquire an attribute of the demonstrator. This acquirable attribute was paying attention or not paying attention to the tape task (that is, good or poor model evaluation). The first hypothesis leads to the prediction that in the Interaction condition, the subjects will predict that they will do better on the tape task when the demonstrator ostensibly performs well than when he ostensibly performs poorly. The mean predictions of performance are presented in Table 1. The data show that the hypothesis is confirmed with respect to this prediction. The first hy-

TABLE 1  
PERFORMANCE AND RATINGS<sup>a</sup> IN EXPERIMENTAL CONDITIONS

Evaluation of model's performance	Interaction condition		No interaction condition		
	Good a	Poor b	Good c	Poor d	
1. Prediction of performance	3.04	3.55	3.45	3.45	a vs. b: 2.309** a vs. c: 1.948* a vs. d: 2.039**
2. Number of errors	5.00	8.00	6.15	7.25	
3. Evaluation of performance	2.33	2.65	2.75	2.00	
4. Desire to know others (not demonstrator)	14.03	14.07	16.05	15.63	
5. Desire to know demonstrator	3.50	3.24	3.85	3.81	
6. Attraction to group	3.48	3.41	3.60	3.50	
7. Interest in tape task	2.90	2.83	3.30	2.70	
y	27	20	20	19	

<sup>a</sup> The higher the rating, the less the quantity indicated (except for Row 2).  
\*  $p < .06$ , two-tailed.  
\*\*  $p < .05$ , two-tailed.

TABLE 2

PERFORMANCE AND RATINGS<sup>a</sup> IN EXPERIMENTAL CONDITIONS FOR SUBJECTS HIGH AND LOW IN DESIRE TO KNOW THE DEMONSTRATOR

IN DESIRE TO KNOW THE DEMONSTRATOR					
Evaluation of model's performance	Interaction condition		No interaction condition		<i>t</i>
	Good a	Poor b	Good c	Poor d	
Subjects High in Desire to Know Demonstrator					
1. Prediction of performance	3.00	3.69	3.78	3.67	a vs. b: 2.060*
2. Number of errors	5.93	7.25	5.38	7.56	
3. Evaluation of performance	2.40	2.62	2.62	2.89	
4. Desire to know others (not demonstrator)	12.33	11.63	14.75	11.78	
5. Attraction to group	3.27	2.94	3.38	3.11	
6. Interest in tape task	2.73	2.69	2.88	2.22	
<i>N</i>	15	16	8	9	
Subjects Low in Desire to Know Demonstrator					
1. Prediction of performance	3.08	3.39	3.50	3.30	a vs. b: 2.161*
2. Number of errors	3.83	8.92	6.67	7.30	
3. Evaluation of performance	2.25	2.69	2.83	2.90	c vs. d: 2.103*
4. Desire to know others (not demonstrator)	10.17	17.08	16.92	19.10	
5. Attraction to group	3.75	4.00	3.75	3.90	
6. Interest in tape task	3.17	3.50	3.58	3.56	
<i>N</i>	12	13	12	10	

<sup>a</sup> The higher the rating, the less the quality rated (except for Row 2).

\*  $p < .05$ , two-tailed.

pothesis also leads to the prediction that the subjects will evaluate their performance (after working on the task) better when the model does better in the Interaction condition. The prediction is not supported by the data in Table 1. The first hypothesis also suggests the corollary that in the No Interaction condition, neither of the predictions made above will be supported. The data of Table 1 support this corollary. However, the most stringent confirmation of this first hypothesis would be to find that the difference between the good and poor Interaction conditions is greater than the difference between the good and poor No Interaction conditions. The significance of the difference was tested by a method suggested by Walker and Lev (1953, p. 158) which is essentially a *t* test of the difference between differences between means. This test was not significant for prediction of performance, or evaluation of performance. In general, however, the results provide some support for this hypothesis with respect to two of the predictions made from it.

The second hypothesis was that in the Interaction condition, the subjects will actu-

ally perform better on the listening task when the model had done well than when he had done poorly. Table 1 shows that this prediction is supported by the number of errors made in each condition. As would be expected from the hypothesis, there is no significant difference between the two No Interaction conditions. However, the Walker-Lev test of the difference between differences between means was not significant. Thus, the study provides some support for the hypothesis that the social influence in the present study influences the subjects' performances on the task.

Also presented in Table 1 are the mean ratings of attraction to the other members of the group, to the demonstrator, and to the group as a whole, and interest in the listening task, all made after the subjects had performed the task. None of the differences among the conditions is significant, nor are the Walker-Lev tests among the means, showing that the experimental conditions tend to influence performance variables significantly without influencing attraction variables.

*Attraction.* The work of Back (1951) and others has suggested that high attraction is



associated with conformity. Subjects who responded to the experimental conditions with high attraction to the demonstrator (model) may therefore have reacted differently from those who responded with low attraction. To test this possibility, the subjects were divided at the median of attraction to the demonstrator, and the experimental differences examined. In Table 2, it can be seen that attraction is not necessarily associated with the subject's conforming to the level of performance of demonstrator. Among subjects highly attracted, only the subjects' predictions of their performances showed the expected condition difference significantly. Subjects of low attraction to the model, on the other hand, showed the expected condition differences in their actual performances, making considerably fewer errors in the Interaction condition when the model had done well than when he had done poorly. It is

quite clear, then, that the processes examined in the present study are different from the processes involved in studies when attraction does lead to conformity. With regard to both these variables, however, Walker-Lev tests of the difference between differences do not yield significant *t* ratios. Therefore, these conclusions must be regarded tentatively.

*Prior acquaintance.* The experimental results would reasonably be expected to have been influenced by the amount of prior acquaintance among the subjects. Table 3 shows the results separately for those subjects with some and with no prior acquaintance with other subjects. The experimental hypotheses are supported primarily by subjects with no prior acquaintance with other subjects. These subjects in the Interaction condition predict better performances, actually do better, and evaluate their performances better when the demonstrator performs better

TABLE 3  
PERFORMANCE AND RATINGS<sup>a</sup> IN EXPERIMENTAL CONDITIONS FOR SUBJECTS WITH SOME AND NO PRIOR ACQUAINTANCE WITH OTHER SUBJECTS

PERFORMANCE AND RATINGS <sup>a</sup> IN EXPERIMENT 1 NO PRIOR ACQUAINTANCE WITH OTHER SUBJECTS					
Evaluation of model's performance	Interaction condition		No interaction condition		<i>t</i>
	Good a	Poor b	Good c	Poor d	
Subjects with No Prior Acquaintance					
1. Prediction of performance	2.81	3.45	3.25	3.64	a vs. b: 2.262**
2. Number of errors	3.69	8.65	8.33	5.86	a vs. b: 2.998*** a vs. c: 2.134** a vs. b: 2.344** a vs. c: 1.887*
3. Evaluation of performance	2.12	2.80	2.67	2.64	
4. Desire to know others (not demonstrator)	15.75	14.90	15.83	15.85	
5. Desire to know demonstrator	3.50	3.25	3.83	4.00	
6. Attraction to group	3.38	3.85	3.50	3.36	a vs. c: 2.926***
7. Interest in tape task	2.56	2.85	3.75	2.92	
<i>N</i>	16	20	12	14	
Subjects with Some Prior Acquaintance					
1. Prediction of performance	3.36	3.78	3.75	3.00	a vs. c: 2.051*
2. Number of errors	6.91	6.55	2.88	10.50	c vs. d: 1.906* b vs. d: 2.446**
3. Evaluation of performance	2.64	2.33	2.87	3.50	
4. Desire to know others (not demonstrator)	13.73	13.33	16.38	15.17	
5. Desire to know demonstrator	3.73	3.22	3.88	3.50	a vs. b: 2.351** b vs. d: 2.572**
6. Attraction to group	3.64	2.44	3.75	3.83	
7. Interest in tape task	3.45	2.78	2.62	2.50	
<i>N</i>	11	9	8	6	

<sup>a</sup> The higher the rating, the less the quality rated (except for Row 2).  
\*  $p < .10$ , two-tailed.  
\*\*  $p < .05$ , two-tailed.  
\*\*\*  $p < .01$ , two-tailed.

TABLE 4

PERFORMANCE AND RATINGS<sup>a</sup> IN INTERACTION CONDITIONS FOR SUBJECTS HIGH AND LOW IN NUMBER OF NAMES SENT AND RECEIVED

Evaluation of model's performance	High number		Low number		<i>t</i>
	Good a	Poor b	Good c	Poor d	
1. Prediction of performance	3.68	3.62	3.00	3.50	<i>c vs. d</i> : 2.275*
2. Number of errors	5.39	6.46	4.64	9.25	
3. Evaluation of performance	2.15	2.46	2.50	2.81	
4. Desire to know others (not demonstrator)	14.69	14.85	15.14	13.44	
5. Desire to know demonstrator	3.77	3.15	3.43	3.31	
6. Attraction to group	3.77	3.15	3.21	3.62	
7. Interest in tape task	2.92	2.54	2.93	3.06	
<i>N</i>	13	13	14	16	

<sup>a</sup> The higher the rating, the less the quality rated (except for Row 2).\*  $p < .05$ , two-tailed.

than when he performs worse. In fact, these effects are so strong that subjects in the Interaction condition with a good model perform better and evaluate themselves better than those in the No Interaction condition with a good model. On the other hand, subjects who have at least some prior acquaintance with the other subjects behave differently. They do not manifest to a significant degree any of the differences predicted between the good and poor Interaction conditions. In fact, a Walker-Lev test shows that the subjects of no prior acquaintance show these differences to a greater degree. For the number of errors, the Walker-Lev  $t$  ratio is 1.738 ( $p < .10$ , one-tailed), while it is 1.740 for evaluation of performance ( $p < .10$ , one-tailed).

Futhermore, in the No Interaction conditions, the subjects of some prior acquaintance react in a completely unexpected manner. In these conditions the subjects perform better when the model does better; and the mean number of errors is significantly less in the No Interaction condition with a good demonstrator than in the Interaction condition with a good one. Moreover, a Walker-Lev test yields a  $t$  ratio of 2.655 ( $p < .02$ , two-tailed), showing that the subjects of some prior acquaintance in the No Interaction condition do react differently from those of none. The authors confess that, without building a pseudological structure of tenuous assumptions and derivations, they are at a loss to explain the behavior of the persons who had some prior acquaintance.

*Number of names sent and received.* In the actual administration of the Interaction condition, there were two factors preventing all of the subjects from exchanging as many names with the to-be demonstrator as the procedure called for. These factors were: slowness of assembling names by particular demonstrators and subjects; errors by the demonstrator or subjects in sending names. Although these errors were minimal, the subjects in the Interaction condition could still be divided into two groups according to the number of names sent and received. It would be expected that those sending and receiving a high number of names would confirm the hypotheses better than those sending and receiving a low number. The results shown in Table 4 do not support this expectation. With respect to actual performance, the subjects with a low number of names sent and received behaved in line with the model's performance, while the higher did not. However, a Walker-Lev test does not demonstrate a significant difference between the two groups of subjects.

Nevertheless, the surprising direction of the difference does require discussion. A possible explanation is that the subjects with a low number of exchanges of syllables with the demonstrator were still attempting to achieve a certain quality of relationship with him, which would be achieved by making one-self similar to him. On the other hand, those with a high number of exchanges had achieved a satisfactory level of exchange with him and,

therefore, did not have to behave like him to achieve a certain quality of relationship.

*Self-esteem.* The third hypothesis predicted that subjects low in self-esteem would be more responsive to the experimental differences in the expected ways than would subjects high in self-esteem. The results shown in Table 5 indicate that exactly the opposite tended to occur. The subjects with low self-esteem evidenced no significant differences at all between experimental conditions on any dependent variable. On the other hand, in the Interaction condition, subjects high in self-esteem predicted that they would perform better, actually performed better, and evaluated their performance better when the demonstrator did well than when he did poorly. No differences occurred for these subjects in the No Interaction condition. In fact, the subjects in the Interaction condition with a good demonstrator predicted better performance than those in the No Interaction condition with the same quality of demonstrator.

A Walker-Lev test of the significance of the differences in the two Interaction conditions between the subjects high and low on self-esteem yields a ratio of 1.913 ( $p < .06$ , two-tailed) for evaluation of performance. Parallel tests for prediction of performance and number of errors do not yield significant  $t$  ratios.

The obtained pattern of results shows that the overall result reported in Table 1 for the whole sample was generated more by the subjects high in self-esteem than by those lower in self-esteem. These results are that the subjects perform better in the Interaction condition with a good demonstrator. This may have been produced by the high self-esteem subjects' reacting to the experimental conditions by high self-evaluation. However, without the appropriate social situation, the high self-esteem persons do not behave in this fashion, since they did not show any difference between the two No Interaction conditions. This tendency of persons with high self-esteem to react to social situations

TABLE 5  
PERFORMANCE AND RATINGS<sup>a</sup> IN EXPERIMENTAL CONDITIONS FOR SUBJECTS  
HIGH AND LOW IN SELF-ESTEEM

PERFORMANCE AND RATINGS HIGH AND LOW IN SELF-ESTEEM					
Evaluation of model's performance	Interaction condition		No interaction condition		<i>t</i>
	Good a	Poor b	Good c	Poor d	
Subjects High in Self-Esteem					
1. Prediction of performance	2.80	3.64	3.14	3.62	a vs. b: 2.509**
2. Number of errors	3.33	7.77	7.86	7.54	a vs. b: 4.976***
3. Evaluation of performance	1.87	2.64	2.71	2.85	a vs. b: 2.587** a vs. c: 2.211*
4. Desire to know others (not demonstrator)	15.27	13.21	15.86	15.25	
5. Desire to know demonstrator	3.60	3.21	3.86	3.75	
6. Attraction to group	3.40	3.57	3.14	3.89	
7. Interest in tape task	2.53	2.43	3.14	2.77	
<i>N</i>	15	14	7	13	
Subjects Low in Self-Esteem					
1. Prediction of performance	3.30	3.47	3.67	3.14	
2. Number of errors	6.50	8.27	5.50	6.71	
3. Evaluation of performance	2.80	2.67	2.83	3.00	
4. Desire to know others (not demonstrator)	14.00	14.87	16.83	16.29	
5. Desire to know demonstrator	3.30	3.27	4.00	4.00	
6. Attraction to group	3.60	3.27	3.83	3.71	
7. Interest in tape task	3.20	3.70	3.42	2.83	
<i>N</i>	10	15	12	7	

<sup>a</sup> The higher the rating, the less the quality rated (except for Row 2).  
\*  $p < .05$ , two-tailed.  
\*\*  $p < .02$ , two-tailed.  
\*\*\*  $p < .001$ , two-tailed.



TABLE 6  
PERFORMANCE AND RATINGS<sup>a</sup> IN EXPERIMENTAL CONDITIONS FOR FIRST BORN AND ONLY CHILDREN AND FOR LATER BORN

ONLY CHILDREN AND FOR LATER BORN						
Evaluation of model's performance	Interaction condition		No interaction condition		<i>t</i>	<i>p</i> <sup>b</sup>
	Good a	Poor b	Good c	Poor d		
First and Only Children						
1. Prediction of performance	3.20	3.64	3.29	3.31	c vs. d: 2.09	<.05
2. Number of errors	5.67	5.65	5.00	8.62		
3. Evaluation of performance	2.33	2.54	2.43	3.08		
4. Desire to know others (not demonstrator)	14.73	12.43	15.71	16.08		
5. Desire to know demonstrator	3.93	3.14	3.86	4.00		
6. Attraction to group	3.33	2.86	3.43	3.00		
7. Interest in tape task	3.47	3.14	3.64	3.62		
<i>N</i>	15	14	14	12		
Later Borns						
1. Prediction of performance	2.83	3.46	3.80	3.67	a vs. b: 1.839	<.10
2. Number of errors	4.17	10.85	9.20	5.17	a vs. c: 2.173	<.05
3. Evaluation of performance	2.33	2.85	3.40	2.50	a vs. b: 2.819	<.02
4. Desire to know others (not demonstrator)	15.17	15.69	16.40	14.00	b vs. d: 1.988	<.07
5. Desire to know demonstrator	3.17	3.31	3.60	3.33	a vs. c: 2.246	<.05
6. Attraction to group	3.50	3.92	3.60	3.17	c vs. d: 1.965	<.08
7. Interest in tape task	2.42	2.92	3.00	2.67		
<i>N</i>	12	13	5	6		

<sup>a</sup> The higher the rating, the less the quality indicated (except for Row 2).

<sup>b</sup> All *p* values two-tailed.

mainly when the consequences of the reaction would be ego enhancing has also been found by Burnstein, Stotland, and Zander (1961), and Stotland, Thorley, Thomas, Cohen, and Zander (1957), and has been discussed by Cohen (1959). These persons evidently maintain their high self-esteem through mechanisms similar to the one manifested here. Persons low in self-esteem may react more readily in social situations in which social pressure is exerted in the form of an explicit request, demand, or expectation by other persons. This was the case in Hovland and Janis (1959) and Stotland et al. (1957). In the present study, no requests, demands, or expectations were explicitly communicated to the subjects regarding their behaving like the demonstrator. Therefore, the low self-esteem persons behaved the same in both the Interaction and the No Interaction conditions.

*Birth order.* Schachter (1959) has demon-

strated the significance of birth order for reactions to social situations. To explore this variable further, the subjects were divided into two groups, first and only children, and later born children, and the results tabulated separately for each. Table 6 clearly shows that the later borns supported the experimental hypothesis much more than the first born and only children. The first born and only children show only one significant difference between conditions, which was in their evaluation of performance in the No Interaction conditions. On the other hand, in the Interaction condition, the later borns predicted better performance and actually performed better when the model was good than when he was poor. In the interaction conditions, the Walker-Lev test between the two samples yields a *t* ratio of 2.278 (*p* < .05) for number of errors, showing that the later borns reacted more to the experimental differences. In

fact, in the No Interaction condition there was some tendency for the later borns to react in a manner completely opposite to the way they reacted in the Interaction condition. That is, they evaluated their performance better when the model was worse. Further indications of the reversal are that, when the model was good they made better predictions and evaluations in the Interaction condition than in the No Interaction condition; and that, when the model was poor, they performed worse in the Interaction condition than in the No Interaction condition. The Walker-Lev test, between the two samples in the No Interaction conditions, produces a  $t$  ratio of 1.779 ( $p < .10$ , two-tailed) for number of errors, and 2.719 ( $p < .02$ , two-tailed) for evaluation of performance.

#### SUMMARY

In the present study a condition of high interaction was created in one task. The effects of this interaction on perceived similarity to the high interactor and to his level of performance were tested in another task. The intention was to test for interaction per se without the involvement of the content of the interaction. Twenty groups of six subjects each (all college students) were formed to carry out a word assembly task so arranged that one of the six would be the center of the interaction. Subsequently, the groups were given the task of detecting the number of repetitions of the same sounds in a tape recording in a foreign language. This second part of the experiment was carried out by assigning groups to one of four situations: the high interactor from the first task was the model with public praise for his performance; the high interactor was the model with public criticism of his performance; a subject not central to the interaction in the first task was the model with public praise for performance; and a subject not central to the interaction in the first task was the model with public criticism of his performance. Measures were taken of the subjects' predictions of the performance, of their actual performance, and their evaluation of it.

The findings were:

1. Consistent with the hypothesis of the study, the subjects rated their performances, and actually performed on the counting task, more in line with the interaction model's ostensible level of performance than with the no interaction model's.

2. Subjects low in attraction to the model appeared to behave more consistently with the hypothesis in their actual performances than did subjects high in attraction.

3. Subjects who had no prior acquaintance with other subjects confirmed the hypothesis more than those of some prior acquaintance.

4. Subjects who for a variety of reasons exchanged a low number of words with the interaction model performed on the counting tasks more consistently with his ostensible level of performance than did those who exchanged a high number.

5. Subjects high in self-esteem behaved more consistently with the hypothesis above than did subjects low in self-esteem.

6. Subjects who had been later born children behaved more consistently with the experimental hypothesis than did the first and only children.

#### REFERENCES

- BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9-23.
- BURNSTEIN, E., STOTLAND, E., & ZANDER, A. Similarity to a model and self-evaluation. *J. abnorm. soc. Psychol.*, 1961, **62**, 257-264.
- COHEN, A. R. Some implications of self-esteem for social influence. In C. Hovland & I. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale Univer. Press, 1959. Pp. 102-120.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- HOVLAND, C., & JANIS, I. (Eds.) *Personality and persuasibility*. New Haven: Yale Univer. Press, 1959.
- MICHOTTE, A. E. *La perception de la causalite*. Paris: Vocin, 1946.
- SCHACHTER, S. *The psychology of affiliation*. Stanford, Calif.: Stanford Univer. Press, 1959.
- STOTLAND, E., COTTRELL, N. B., & LAING, G. Group interaction and perceived similarity of members. *J. abnorm. soc. Psychol.*, 1960, **61**, 335-340.
- STOTLAND, E., THORLEY, S., THOMAS, E. J., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55-63.
- WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received January 23, 1961)

## DIMENSIONS OF SUBJECTIVE ADJUSTMENT<sup>1</sup>

JOSEPH VEROFF, SHEILA FELD,<sup>2</sup> AND GERALD GURIN

*Survey Research Center, University of Michigan*

Research programs assessing mental health have a broad range of problems that are possible of exploration. The criteria used to evaluate health or illness are diverse. Problems of social adjustment, aspects of self-esteem, psychosomatic complications, creativity all can be investigated. Or one can simply identify disturbance pragmatically by isolating those admitted to mental hospitals or those who have been exposed to some sort of psychiatric treatment. Recent thinking about this confusing diversity of criteria for evaluating mental health and illness has converged on the reasonable viewpoint that different criteria for such evaluations are not only equally possible and equally valid but are required for a complete appraisal of mental functioning (Eaton, 1951; Fiedler, Dodge, Jones, & Hutchins, 1958; Jahoda, 1950, 1953, 1958; Scott, 1958; Smith, 1950, 1959).

In the present study, a factorial classification of various self-descriptions of adjustment and distress. the data for inferences about mental health were limited to self-reports, thereby ruling out considerations of criteria for mental health that are based on other sources of inference, such as the assessment of actual behaviors or the diagnoses of an "objective" clinical observer or the judgments of peers. But even within this focus on self-evaluations of psychic distress, there exists the possibility of using several different kinds of dimensions for evaluating "good" and "poor" mental health. Self-descriptions are often assumed to be a single criterion to be compared with criteria based on evaluations from psychiatrists, friends, etc. (Eaton, 1951; Fielder et al., 1958). But there are many

facets to self-descriptions. And many ways of defining felt distress appear in the research literature on mental health. Different indices of dysfunctioning focus on different aspects of psychic adjustments: psychosomatic complaints (MacMillan, 1957), ways of perceiving the self (Rogers, 1951), life satisfactions in marriage (Terman & Wallin, 1949), in work (Herzberg, Mausner, & Snyderman, 1959; Morse, 1953), life happiness in general (Inkeles, 1960). Each of these approaches is by itself inadequate for fully identifying distress. Usually, the investigator concentrates on one aspect of these subjective evaluations. In so doing he may be trapped into identifying this aspect of mental health as the most important criterion for defining mental health. In the present study, rather than choosing among these various criteria of distress, a multiple criterion approach was adopted. The research to be reported suggests that multiple criteria are needed in a description of subjective adjustment, and further suggests what these criteria might be psychologically.

The present paper is a further report on a nationwide sample survey recently reported on by the present authors (Gurin, Verooff, & Feld, 1960). A total of 2,460 adult respondents were selected by area sampling probability methods to constitute a representative cross-section of adults, 21 years of age or older, living in private households in the United States (Kish, 1953). This sample omits the hospitalized population. They were interviewed about their lives, their experiences of distress or satisfaction in many *different* kinds of ways. In utilizing a multiple criterion approach to the measurement of feelings of adjustment, each index was considered separately, on its own terms. No one measure was weighted more heavily than the other as a "better" measure of psychic distress. Rather, we assumed that unhappiness, the sense of inadequacy, tension in a marriage or with children, problems at work, constant worries,

<sup>1</sup> The investigations reported in this paper are based on data obtained within a national sample survey supported by the Joint Commission on Mental Illness and Health. Analysis of these data was supported by a grant from the National Institute of Mental Health, United States Public Health Service, Project M2280.

<sup>2</sup> Now at Yale University.



feelings of an impending nervous breakdown. all reflect different areas of potential distress. In the analyses already reported, the information in the survey was examined to see how *each* index varied from social group to group. and how *each* was related to patterns of seeking professional help for personal problems.

A major question remains from these analyses: How are the various indices covering different aspects of life adjustments related one to another—especially for groups for whom all indices were relevant—married women with children and employed married men with children who fall into our non-institutionalized United States population? This is the problem for this paper. In this study, factor analytic techniques were used to assess the communalities in response to the diverse questions asked of the national sample. The emerging factor structure may be indicative of basic psychological factors underlying the determination of self-descriptions of adjustment, and as such it represents an empirical basis for theorizing about the significance of these disparate measures.

The factor analytic technique allows the researcher to look at clusters of responses in different ways. It may very well be in our analysis that certain clusters will reflect accommodations to strong cultural pressures to perceive the self in a prescribed manner. Or it may be that the clusters reflect deeply felt reactions to the common psychological problems faced by most people in American society. In general we will be following the latter interpretation of the significance of the analyses. We recognize that the other interpretation is possible and will highlight the alternate view where it appears appropriate.

We have approached the problem of investigating self-descriptions with the assumption that separate components to self-evaluations were more than likely. Orthogonal factors generating intercorrelations of responses are part of the assumptive framework used in procedures for this study in factor extraction and rotation. This assumption will not guarantee multiple dimensions of self-descriptions, but it favors such a possibility of data analysis.

All the indices of adjustment are based on self-descriptions. As such they raise the question of honesty of response. Some distortion in self-report is inevitable in a study of subjective adjustment. However, we would expect such distortion to be mainly reflected in a generalized tendency to admit or deny distress in the different life-areas investigated in this study, rather than in systematic distortions on certain of the measures. Since a factor analytic study would be mainly affected by systematic rather than generalized distortions, we might expect the error springing from such distortions to be less critical for the analyses reported here than in other analyses—for example, those which would involve interpretations of the absolute values of the responses to the different questions on which the adjustment measures are based. However, some systematic distortion may exist, and as such would add a source of error to the results.

Error springing from the problem of honesty of response may also be introduced when factor structures are compared for different subgroups of the population. For example, in the comparison between men and women that will be discussed in this paper, the question may arise as to whether the sex differences were affected by differences in either the prescribed stereotypes of responding for men or for women or in how willing men and women are to talk to an interviewer (usually a woman) in an interview situation. However, as will be noted in the later discussions, the similarities in the factor structures of men and women are more striking than the differences, and the differences that do exist do not seem relevant to the question of differences in honesty of response.

In the present study we have considered only a portion of the total sample of 2,460 subjects. Three interview forms were used in the study, each assigned randomly to one-third of the subjects. One-third of the total sample of men and two-thirds of the sample of women were asked all the questions dealt with in the present paper; certain questions were applicable only to currently married subjects with children or, among the men, to working subjects. The final *Ns* are 542 women and 255 employed men who are currently married and have children.

PROCEDURE

The expression of distress was measured in six areas: general feelings of distress, attitudes towards the self, marital adjustment, adjustment as a parent,

job adjustment (men only), and psychological and psychosomatic symptom complaints. The separate indices are listed in Table 1 under these headings together with the questions on which each index

TABLE 1  
INDICES USED IN FACTOR ANALYSES OF SELF DESCRIPTIONS OF DISTRESS

Questions	Indices
<b>General Feelings of Distress</b> Everybody has some things he worries about more or less. What kinds of things do you worry about most? Do you worry about such things a lot, or not very much? Taking things all together, how would you say things are these days—would you say you're <i>very happy</i> , <i>pretty happy</i> , or <i>not too happy</i> these days? Compared to your life today, how do you think things will be 5 or 10 years from now—do you think things will be happier for you than they are now, not quite as happy, or what? Have you ever felt that you were going to have a nervous breakdown?	Worrying All the time Sometimes Never Unhappiness Not too happy Pretty happy Very happy Future unhappiness Not as happy About the same Happier Nervous breakdown Yes No
<b>Attitudes toward Self</b> People are the same in many ways, but no two people are exactly alike. What are some of the ways in which you're different from most other people? Same as above. Many people when they think about their children, would like them to be different from themselves in some ways. If you had a son (daughter) how would you like him (her) to be different from you? How about your good points? What would you say were your strongest points?	Lack of uniqueness of self Sees no differences from others Mentions differences Lack of self-acceptance (degree to which subject gives negative evaluation of the self in describing differences from others) Negative Ambivalent Neutral Positive Very positive Shortcomings in the self Wants child to be different Does not want child to be different Lack of strong points in the self Sees no strong points Mentions strong points
<b>Marital Adjustment</b> Many men (women) feel that they're not as good husbands (wives) as they would like to be. Have you ever felt this way? What kinds of things make you feel this way? Do you feel this way a lot of times, or only once in a while? Taking things all together, how would you describe your marriage—would you say your marriage was very happy, a little happier than average, just about average, or not too happy? Even in cases where married people are happy there have often been times in the past when they weren't too happy—when they had problems getting along with each other. Has this been true for you?	Feelings of marital inadequacy A lot of times Once in a while Never Marital unhappiness Not too happy Average A little happier than average Very happy Experience of problems in marriage Mentions problems Never had problems
<b>Adjustment as a Parent</b> First, thinking about a man's (woman's) life, how is a man's (woman's) life changed by having children? Most parents have had some problems in raising their children. What are the main problems you've had in raising your children? Many men (women) feel that they're not as good fathers (mothers) as they would like to be. Have you ever felt this way? (If yes) What kinds of things have made you feel this way? Have you felt this way a lot of times, or only once in a while?	Negative orientation to children: Coder rating of change seen as accompanying children Negative Neutral Positive Experience of problems in raising children Mentions problems Never had problems Feelings of inadequacy as a parent A lot of times Once in a while Never
<b>Psychological and Psychosomatic Symptom Complaints</b> Do you ever have any trouble getting to sleep or staying asleep? Check one: Nearly all the time (1); Pretty often (2); Not very much (3); Never (4). Have you ever been bothered by nervousness, feeling fidgety and tense? Check one: as above. Do you feel you are bothered by all sorts of pains and ailments in different parts of your body? Yes (2); No (4). For the most part, do you feel healthy enough to carry out the things you would like to do? Yes (4); No (2). Do you find it difficult to get up in the morning? Check one: Nearly all the time (1) Pretty often (2); Not very much (3); Never (4). Are you troubled by your hands sweating so that you feel damp and clammy? Check one: Many times (1); Sometimes (2); Hardly ever (3); Never (4).	Symptom Factor 1: Psychological Anxiety <sup>a</sup> Six ranks ranging from very high to very low scores Symptom Factor 2: Physical Health <sup>a</sup> Three ranks Symptom Factor 3: Immobilization <sup>a</sup> Six ranks

TABLE 1—Continued

Questions	Indices
Have you ever been bothered by shortness of breath when you were not exercising or working hard? Check one: Many times (1); Sometimes (2); Hardly ever (3); Never (4).	Symptom Factor 4: Physical Anxiety <sup>a</sup> Six ranks
Have you ever been bothered by your heart beating hard? Check one: as above.	Summary Symptom Score <sup>c</sup> Ten ranks
Summation of 16 symptom questions used in Stirling County study (MacMillan, 1957).	
Job Adjustment Taking into consideration all the things about your job, how satisfied or dissatisfied are you with it?	Job dissatisfaction Dissatisfied Ambivalent Satisfied Very satisfied
Have you ever had any problems with your work—times when you couldn't work or weren't getting along on the job, or didn't know what kind of work you wanted to do?	Experience of problems with job Mentions problems Never had problems Feelings of inadequacy on job Not very good Average Little better than average Very good
How good would you say you are at doing this kind of work—would you say you were very good, a little better than average, just average, or not very good?	

<sup>a</sup> These symptom factor indices are based on prior factor analyses of a symptom list of 20 items which were taken, with slight changes, from the Health Opinion Survey (MacMillan, 1957). Each index represents the two items most clearly typifying the factor; these items are ones having not only high loadings on the factors they are assumed to represent, but also minimal loadings on other factors.

<sup>b</sup> Scores on the symptom factors are the sum of the values in parentheses.

<sup>c</sup> For a list of these items see Gurin, Veroff, and Feld (1960, Table 7).

was based. For a full description of the coding procedures for each index and the complete interview schedule the reader is referred to *Americans View Their Mental Health* (Gurin et al., 1960).

For purposes of intercorrelation, the responses to the questions constituting each index were ranked—the questions constituting each index were ranked with a rank of 1 indicative of highest distress. As can be seen in Table 1, the number of rank positions varied from index to index, with seven of the indices having only two rank positions.

Separate matrices of intercorrelations of these indices were obtained for men and women. Since the sex differences in response to some of the separate indices were very great, it was felt that the possibility of distinctive factor structures emerging for men and women warranted consideration. (Separate factor analyses for different age and educational groups may also be necessary; we anticipate following through these possibilities in future analyses.)

Kendall's tau beta was computed for each pair of indices. This particular statistic was used for two reasons. It avoids the necessity of collapsing categories of coding, as would have to be done if we employed tetrachoric correlations or phi coefficients; it also avoids the necessity of making unwarranted assumptions about units of measurement, as would be required if Pearsonian product-moment correlations were used. The logic of factor analysis is applicable to tau beta's since the statistic has been found to operate similarly to Pearsonian product-moment correlations usually used in factor analyses; the sampling errors and distributions of the two statistics are similar.<sup>3</sup>

<sup>3</sup> The authors would like to express their appreciation to William L. Hays and to John A. Sonquist for their help with these statistical problems and procedures.

The centroid method of factor extraction and IBM 650 Principle Axis Varimax program for rotation were used to obtain the final factor structures for men and women.

## RESULTS AND DISCUSSION

Tables 2 and 3 present the matrices of intercorrelations for men and women, respectively. The significance of tau beta can be estimated: for the men ( $N = 255$ ), tau's needed for significance at the .05 and .01 levels of confidence are .082 and .108, respectively; for the women ( $N = 542$ ), the comparable figures are tau's of .054 and .091, respectively (cf. Kendall, 1948). Since the size of the samples involved here is large, and the size of the correlations required for significant deviations from zero is therefore relatively small, our major interest in these data is in the communalities of responses shown in the factor analysis.

In Table 4, the rotated factor loadings for both men and women are presented. The order in which the factors were extracted for men was 1,2,4,5,3; for women it was 1,3,2,4,5. In attempting to interpret the factor structures for men and women, we limited our consideration to those items which had factor loadings of .30 or higher.

In the main, the factor structures for men and women are similar; four parallel factors emerge and one additional factor appears for men. The similarity in these factor structures



for men and women is an important finding; unexplained sex differences in obtained relationships are often the bane of the psychological researcher. However, the differences between men and women in the factor loadings of the various items on each factor are provocative. We highlight these in the discussion to follow.

### Factor Structure for Men

**Factor 1: Felt psychological disturbance.** The variables that load highly on this factor are derived from the symptom checklist and questions concerning psychological and psychosomatic complaints (*psychological anxiety, immobilization, physical anxiety, and summary symptom score*). This factor represents an admission to symptoms that have particular implications for potential psychological dysfunctioning.

Since the variables that load highly on this factor all derive from a symptom checklist, filled out by the respondent himself, the question may arise as to what extent this factor

represents a response set. It may be noted, in this connection, that the physical ill health symptom index, which also derives from this symptom checklist, is *not* highly loaded on this factor, which argues against a general "response set" interpretation.

It is noteworthy that this factor, entirely composed of symptom factor indices, does not have high loadings from other measures and further, that the symptom factor indices do not have high loadings on other factors, except on Factor 5. This, as we shall see, raises certain questions about the diagnostic use that has been made of symptom checklists.

**Factor 2: Unhappiness.** The general evaluations of *unhappiness* and *marital unhappiness* are the most highly loaded indices on this factor. In addition, the *problems with children* and *shortcomings* indices are also loaded on this factor. Thus "unhappiness" among men appears to be a general factor, tied to distress in several central life-areas—marriage, children, and the self-concept. While the loading of these different indices on

TABLE 3  
INTERCORRELATIONS (tau beta) OF 19 INDICES OF SUBJECTIVE ADJUSTMENT  
(N = 542 married women with children<sup>a</sup>)

	Worrying	Unhappiness	Future unhappiness	Nervous breakdown	Lack of uniqueness of self	Lack of self-acceptance	Shortcomings	Lack of strong points	Marital inadequacy	Marital unhappiness	Marriage problems	Negative orientation to children	Problems with children	Inadequacy as parent	Psychological anxiety	Physical ill health	Immobilization	Physical anxiety	Summary symptom score
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
01																			
02																			
03		+18 <sup>b</sup>																	
04			-05																
05			-04																
06				16	-04	10	05	-01	08	05	02	06	04	19	06	06	08	14	
07				11	02	02	00	04	-01	44	24	06	05	03	19	10	02	15	18
08			-04		05	-03	00	08	-12	01	-02	-03	-09	-02	-06	01	-10	05	01
09				-01	-07	13	08	01	14	16	18	-01	-05	04	-06	00	-01	-04	-03
10						-04	-05	13	-01	02	-04	-00	08	12	11	03	07	10	08
11							07	07	14	02	01	-00	07	16	08	-03	14	07	08
12								-03	20	04	10	03	-06	-01	01	07	-04	03	-01
13									01	04	00	-03	15	37	15	-00	16	14	16
14										04	14	03	07	04	15	06	01	18	20
15											28	07	07	13	18	-06	08	14	19
16												08	04	-01	01	-04	07	08	07
17														17	14	-01	03	06	10
18															16	00	15	13	16
19																24	23	32	58
																	07	25	38
																		22	33
																			58

<sup>a</sup> Includes only women responding to the two forms of the interviews on which all 19 items were asked. *N*s for specific correlations vary slightly due to incomplete data for some subjects.  
<sup>b</sup> Plus sign is omitted from all other figures in table. All correlations are positive unless otherwise noted.

TABLE 4  
FACTOR LOADINGS FROM NORMALIZED VARIMAX ROTATIONS<sup>a</sup>

Indices of subjective adjustment	Factor 1		Factor 2		Factor 3		Factor 4		Factor 5		h <sup>2b</sup>	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Worrying	+18 <sup>c</sup>	13	28	-10	-08	-02	-26	-07	07	-27	19	11
Unhappiness	14	11	56	-66	04	07	09	01	-08	-14	35	47
Future unhappiness	00	02	08	00	13	08	-11	28	-32	07	14	09
Nervous breakdown	29	41	02	-13	39	16	12	06	19	-35	29	34
Lack of uniqueness of self	-02	-03	04	02	10	-02	31	33	04	09	11	12
Lack of self-acceptance	26	08	03	02	07	-18	40	03	00	-22	23	09
Shortcomings	00	05	31	-03	-28	-34	-07	-06	08	-02	19	12
Lack of strong points	06	02	18	02	-05	-01	31	34	02	-06	13	12
Marital inadequacy	28	10	12	01	-37	-58	09	-05	21	-11	28	36
Marital unhappiness	10	11	54	-64	07	-01	19	06	-09	05	35	43
Marriage problems	21	07	20	-42	-22	-26	21	-06	14	00	20	25
Negative orientation to children	04	05	06	-14	12	-05	-05	-19	23	15	08	08
Problems with children	-03	02	31	-08	-16	-22	-06	18	-05	-27	13	16
Inadequacy as parent	13	08	11	-06	-59	-56	05	05	-02	-09	38	33
Psychological anxiety	70	57	24	-14	-09	-11	-19	-10	-05	-28	50	44
Physical ill health	18	46	08	02	-05	14	06	12	-55	-15	35	27
Immobilization	53	36	04	01	-17	26	02	-14	02	06	31	23
Physical anxiety	42	62	07	-19	-17	-15	07	03	-49	01	45	44
Summary symptom score	66	85	14	-14	-19	-12	00	00	-37	-06	64	76
Job dissatisfaction	15		20		-09		14		02		09	
Work problems	11		01		-44		01		-10		22	
Job inadequacy	08		-02		-15		28		20		15	

<sup>a</sup> Based on samples of 255 men and 542 women.

<sup>b</sup> h<sup>2</sup> is equal to the communality of the item.

Plus sign is omitted from all other figures in table. All factor loadings are positive unless otherwise noted.

an unhappiness factor is not surprising, it is interesting to note that a quite different pattern appears for women where, as we will see, only the marital indices are loaded on the unhappiness factor.

**Factor 3: Social inadequacy.** This factor seems to represent the self-questioning that goes on when a man considers himself in relationship to his social world. Some men entertain considerable self-doubt in their role relationships with others; others are self-confident or refuse to admit to doubt. Items loaded on this factor seem to be measuring the extent to which a man asks himself: Am I doing a good job at the things I am doing—worker, husband, father? Such feelings are objectified into *job problems* (.44), *inadequacy as a parent* (.59), and *marital inadequacy* (.37).

If this factor is viewed as measuring general social inadequacy, the relative weights of the factor loadings of these indices suggest that feelings of inadequacy as a father may be the best indications of a man's general feelings of inadequacy, or at least may be better measures of this factor than feelings of inadequacy as a husband or as a worker.

The difference between the loadings of the parental and marital indices is particularly interesting. It suggests that admission to inadequacy may involve minor rather than major deviance from ideal role behavior more often when one is thinking of one's behavior as a husband than when one is thinking of one's behavior as a father. In marriage—perhaps because of the greater complexity of the role—we find built-in allowances permitting deviance for men. This is rather graphically illustrated in the almost jocular sanction given to minor inattentiveness on the part of the husband to his wife—forgetting anniversaries, not helping with the dishes, going out with the boys. Being a father does not seem to provide such culturally approved safety valves, perhaps because the demands of the role are less complex—supporting the child being such an all important feature of the role. Consequently admission to inadequacy as a father may be diagnostic of more serious feelings of social inadequacy than admission to inadequacy as a husband.

Also loaded on this factor is the experience, some time in one's life of having had or been

close to a "nervous breakdown." For a man, a nervous breakdown may be a crucial indication of collapse in the face of pressures in the outside world, an extreme disruption in his ability to face his life obligations.

In this factor, as in the others, the investigation of mental health through the layman's self-perceptions has yielded a cluster of indices that has a parallel in the "experts' attempts to establish criteria for mental health. One of the six major criteria for mental health that Jahoda (1958) describes is that of "environmental mastery." Included in this category is the person's adequacy in a variety of life situations—love, work, play, interpersonal relations. While this criterion is usually evaluated from the "objective" observer's point of view, the present results indicate it is an important aspect of self-evaluations as well.

*Factor 4: Lack of identity (versus heightened self-awareness).* In the two factor analyses this is the only factor which emerges in a bipolar fashion (that is, with high loadings in both the negative and positive direction). On the one pole are found *lack of uniqueness of self*, together with *lack of strong points in the self*; on the other pole we find *lack of self-acceptance*. This at first seems like a perplexing factor, but results stemming from our previous analysis of self-perception in the national survey, suggest a possible explanation of the factor (Gurin et al., 1960, Ch. 3). As we noted in that study, responses to questioning about the self are influenced both by how introspective a person is and by how positively or negatively he evaluates himself. Furthermore, there was some evidence that these two determinants are interrelated, that tendencies toward introspection and self-awareness are related to feeling negative about the self. Being able to see negative traits in the self implies a certain degree of introspection.

In the light of these previous findings, we interpret one pole of this factor to be this heightened self-awareness, negatively loaded primarily with the lack of self-acceptance index. Lack of self-acceptance was coded when the respondent reported any negative aspects of the self, something which, as we have indicated, we interpreted as a particular in-

dication of self-awareness. On the other pole we have an opposite reaction—a rejection of looking into the self, a rejection of considering the self as an object of appraisal, what we might call a lack of identity. The indices positively loaded on this pole of the factor are lack of strong points in the self and lack of uniqueness of self (for example, "There is nothing good about me," "I'm not different from anyone else"). Lack of strong points, in short, at least in its loading on this factor, is not so much a "negative" perception of the self as it is an indication of an inability to introspect.

This factor is of particular interest because theoretical discussions of positive mental health have tended to link self-awareness and positive feelings about the self: "a healthy person knows who he is and does not feel basic doubts about his inner identity" (Jahoda, 1958, p. 29). Without denying this, it is also true (as Jahoda has also pointed out) that these "healthy" self-attitudes at the same time include the ability to see and realistically accept the negative aspects of the self. And it is difficult at times to determine when the "healthy" relationship between self-awareness and realistic self-appraisal turns into a relationship between self-awareness and a negative view of the self. This was illustrated in the previous analyses of the data in this study, where, in comparing the responses to the self-percept questions among the different demographic subgroups of the population, it was noted that "... the social and cultural conditions that lead to introspection and a heightened self-awareness may also lead not only to realistic self-criticism, but to a generally negative self-percept" (Gurin et al., 1960, p. 83).

*Factor 5: Physical distress.* Some of the indices based on the symptom checklist again stand out in this factor—*physical ill health*, *physical anxiety*, and *summary symptom score*. This factor thus seems to be defined by physical dysfunctioning. An important sidelight here is that the highest loading that the *future unhappiness* index has is on this factor (.32). A man's optimism or pessimism about his future is tied in large measure to his potential as a worker and economic provider. Physical disability in a man is a major con-



dition interfering with this capacity. Thus it is reasonable that anticipation of future happiness or unhappiness might draw its heaviest loading on this factor for men.

A comparison between Factors 1 and 5 is of interest because of the considerable overlap in the indices loaded on them. These two factors account for the major portion of the variance of the indices based on the symptom checklist. Physical anxiety and summary symptom score are highly loaded on both factors. In contrast, the two most psychological symptom indices—psychological anxiety and immobilization—appear only on Factor 1, and the most physical symptom index—physical ill health—appears only on Factor 5. This points to a general as well as two specific components of the symptom checklist for men. Some implications of this distinction may be clarified when we examine the factor structure for women, where we will see quite a different structure for these symptoms.

#### *Factor Structure for Women*

Five factors were also extracted for women. Except for one of them, there seem to be some direct parallels between the two factor structures. Although there are differences which suggest ways of conceptualizing male and female differences in psychological functioning, the most striking over-all impression is that the two sets of data are very much alike. The four similar factors in the factor structure of both men and women are: felt psychological disturbance, unhappiness, social inadequacy, and lack of identity. There is no factor in the structure for women comparable to the physical distress factor found in the data from men's responses, and the fifth factor for women is difficult to interpret; only one index loads highly on it—*nervous breakdown*.

In what ways do the factors for females seem to differ from males? Let us consider each in order. There are many different theoretically important dimensions underlying sex as an analytic variable and each might have implications in the present analysis. However, we will select interpretations of differences that seem most plausible but recognize that other interpretations are possible.

*Factor 1: Felt psychological disturbance.* This factor is both similar and dissimilar to Factor 1 for men. Four indices are highly loaded on this factor for both men and women: psychological anxiety, immobilization, physical anxiety, and summary symptom score. But the physical ill health index is highly loaded on this factor for females, whereas for males it was part of a physical distress factor (Factor 5), which we do not find in the factor structure for the females. Furthermore, for men physical anxiety and summary symptom score are not uniquely loaded on Factor 1, as they are for women. Thus it seems that women do not separate admission to bodily complaints from admission to complaints about their general psychological condition; men evidently make this distinction. Perhaps these results are tied to the greater tendencies among women to introspect and to see their problems in psychological terms, in contrast to men's greater tendencies to externalize their problems.

*Factor 2: Unhappiness.* As was observed in the discussion of this factor for men, general unhappiness and marital unhappiness seem to go together, both being highly loaded on this "unhappiness" factor. However, certain differences between men and women also appear. Among women this factor does not have high loadings on two indices—shortcomings and problems with children—that were highly loaded on this factor for men. And among women this factor does have a high loading on one index—*marital problems*—which did not appear among the indices with high loadings on this factor for men.

Among women, then, the unhappiness factor, with high loadings on general unhappiness, marital unhappiness, and marriage problems, seems more exclusively tied to marriage than it was for men. This may be taken as some support for the hypothesis that marriage is more central to a woman's life, and hence plays a more significant role in her general happiness.

But why are not the shortcomings and problems with children indices also included in this factor? With respect to the shortcomings index particularly, one might have expected that marriage, being more central in a woman's life, would be more rather than

less tied to self-evaluation and self-criticism, and that for women a factor heavily loaded on marital unhappiness would show higher rather than lower loadings on a shortcomings index. But one of the findings reported in the larger study may suggest why this is not so. In the analysis of adjustment to the marital role, it was noted that women blame their husbands for marital difficulties more often than men blame their wives (Gurin et al., 1960). This difference in attribution of blame seemed to relate to the male-female role distinctions in marriage. The fact that the woman's role in marriage is more passive and receptive than the man's apparently more often enables her to separate marital failure from failure in the self.

The activity-passivity dimension in role behavior may also help explain why marital unhappiness in women also seems to be separate from distress over the children. Only in this case it is the active role that a woman plays as a mother and the contrast with the more receptive and passive role as a wife that might explain why problems in both areas are not loaded on the same factor, why a woman can apparently separate her reactions to marriage from her response to her children. Unhappiness in the more passive role as wife may actually lead to an overly positive involvement in the more active role as mother—a phenomenon which has often been observed in women who, unhappy and feeling helpless in a frustrating unfulfilling marriage, attempt to compensate by an over-investment in their children.

*Factor 3: Social inadequacy.* The hypothesis that the marriage is more important for the feelings of adjustment for women than for men is given further support in a sex difference obtained in the social inadequacy factor. The loading on this factor for marital inadequacy is .58 for women and .37 for men. Both men and women have very high loadings on the inadequacy as parent index. One further kind of self-questioning, the shortcomings index, is also highly loaded (.31) on this factor for women; this is again similar to the finding for men, where this index falls just short (.28) of our criterion for consideration.

Being a mother and being a wife are both central to a married woman's identity; feel-

ing inadequate in either role is equally important in a woman's feelings of inadequacy and failure. Neither role apparently provides the culturally allowed safety valves for deviance that were noted earlier in relation to the husband's role.

*Factor 4: Lack of identity.* Only one pole of the comparable factor for men emerged for women: Lack of uniqueness of self and lack of strong points in the self again emerged together in this factor, but lack of self-acceptance does not load on the other pole of the factor for women. What might be suggested is that women generally tend to be more introspective, so that a heightened self-awareness is not a special factor for them.

#### *Implications of the Factor Analysis*

The multiple criterion approach to the delineation of a person's adjustment that has received increasing support in recent writings is supported by these factor analyses. Even in that limited evaluation of mental health that is available from what a person tells us as he describes his own adjustment, whether a person gives an "adjusted" response or a "maladjusted" response depends on what the question is. Evidently people do not respond to a series of questions covering life adjustment with one general evaluative framework which one could call the person's subjective adjustment. Rather, a person seems to respond to these series of questions from a number of different frameworks, perhaps represented by the factors that emerged from the factor analysis. There are undoubtedly some people who approach all of the questions about adjustment in a general "distressed" way, for whom being high in one of these factors implies being high in all. In his work with people who seek psychological help the clinician perhaps most often encounters this general discontent that transcends all dimensions of self-description—but this pattern of self-evaluation may be less common among the general population.

It is particularly interesting to observe that the distinctiveness of the factors applies to the symptom checklist as well as to the other adjustment measures that were studied. Since symptom checklists such as the one on which the present symptom indices are based, are often used as a quick psychiatric screening



technique, one might have expected the symptom indices to represent a general factor, highly loaded on all the factors. But this is not so. For both men and women one factor (Factor 1) is composed entirely of symptom indices. And although these symptom items do have some loadings on the other factors, these loadings are not high for women, and for men are high only on Factor 5, which is itself largely composed of the physically loaded symptom indices. Apparently, the major aspects of distress included in a symptom checklist, although partial components of other kinds of psychological distress experiences, are fairly distinct from them. Implicitly, then, the judgments that are based on utilizing such symptom batteries as a diagnostic tool seem to be minimizing the importance of other aspects of feelings of maladjustment. Essentially, they single out one from a number of possible mental health criteria.

This reliance upon one single criterion—whether of symptoms or other factors—has special significance when attempts are made to compare different subgroups of the population in mental illness or health. Such a comparison is questionable if these subgroups vary in their tendencies to experience distress as symptoms or in other ways. If subgroups do vary in the ways in which they experience distress, then the “mental health” evaluation of the different groups would vary according to the particular component of mental health that was chosen as the basis for evaluation; for example, a “symptom” criterion would provide a different ranking of groups than would be obtained from a “felt inadequacy” criterion.

That this does occur has been demonstrated in prior results reported by Gurin et al. (1960). Those results were based on the separate questions used in the present factor analysis. The general point can be illustrated better now that the results of the factor analyses are available. A systematic investigation of subgroup differences in responses to the indices based on the several factors is contemplated. For the present, we can illustrate this problem by considering the responses of groups differing in education.

In order to facilitate these comparisons, factor scores were computed for each of the

factors for men and women. The factor scores were based on summations of those indices that were both highly loaded (.30 or more) on the factor in question *and* were not highly loaded on any other factor. These criteria were established in order to yield factor scores that would both be clearly representative of the respective factors and would be as independent of one another as possible.<sup>4</sup> In the summations, each index was given equal weighting.

The indices comprising each factor score are listed below:

*Factor 1.* For men: psychological anxiety and immobilization; for women: summary symptom score.<sup>5</sup>

*Factor 2.* For men: unhappiness, shortcomings, marital unhappiness, and problems with children; for women: unhappiness, marital unhappiness, and marriage problems.

*Factor 3.* For men: nervous breakdown, marital inadequacy, inadequacy as parent, and job problems; for women: shortcomings, marital inadequacy, and inadequacy as parent.

*Factor 4.* For men: lack of uniqueness, lack of self-acceptance,<sup>6</sup> and lack of strong points; for women: lack of uniqueness and lack of strong points.

*Factor 5.* For men: future unhappiness and physical ill health.

The relationships between education and these factor scores are presented in Table 5. The expectation, based on previous results, that the relationship between education and the experience of distress would vary for the

<sup>4</sup> Since no indices were loaded exclusively on only one factor, complete independence of factor scores was not obtained. But among the 16 possible inter-correlations (10 among the five factors for men and 6 among the four factors for women) 12 yielded tau's of .10 or less. The four exceptions were tau's of .13 between Factors 1 and 2 for men, .24 between Factors 1 and 3 for men, .25 between Factors 1 and 2 for women, and .20 between Factors 1 and 3 for women.

<sup>5</sup> The summary symptom score was used as the sole basis for computing the Factor 1 score for women both because of its extremely high loading on the factor (.85) and because most of the items comprising the other symptom indices loaded on this factor are also included in the summary symptom score.

<sup>6</sup> Because of the bipolar nature of this factor, the rankings for the lack of self-acceptance index presented in Table 1 were reversed in computing the factor score.



TABLE 5  
EDUCATION AND FACTOR SCORES BASED ON SELF-EVALUATIONS OF DISTRESS  
(within-sex)

Factor	Scores <sup>a</sup>	Men			Women		
		Grade school <sup>b</sup>	High school	College	Grade school	High school	College
1. Felt psychological disturbance	H	18 <sup>c</sup>	16 <sup>c</sup>	23 <sup>c</sup>	35 <sup>c</sup>	21 <sup>c</sup>	13 <sup>c</sup>
	MH	21	19	25	25	23	30
	ML	22	24	19	18	28	23
	L	30	41	33	22	28	34
	N	85	112	57	139	315	82
2. Unhappiness	H	35 <sup>c</sup>	29 <sup>c</sup>	28 <sup>c</sup>	32 <sup>c</sup>	17 <sup>c</sup>	17 <sup>c</sup>
	MH	33	24	20	32	27	22
	ML	13	19	28	22	29	33
	L	19	28	24	14	27	28
	N	70	97	50	139	317	82
3. Social inadequacy	H	19 <sup>c</sup>	22 <sup>c</sup>	38 <sup>c</sup>	13 <sup>c</sup>	21 <sup>c</sup>	24 <sup>c</sup>
	MH	19	25	30	19	26	35
	ML	35	24	18	24	25	26
	L	26	29	14	44	28	15
	N	68	96	50	109	263	74
4. Lack of identity	H	15 <sup>c</sup>	6 <sup>c</sup>	2 <sup>c</sup>	7 <sup>c</sup>	5 <sup>c</sup>	7 <sup>c</sup>
	MH	41	34	32	{ 36	{ 31	{ 34
	ML	27	45	40	57	64	59
	L	17	15	26			
	N	84	111	57	138	312	83
5. Physical distress	H	38 <sup>c</sup>	11 <sup>c</sup>	10 <sup>c</sup>			
	M	35	31	39			
	L	27	58	51			
	N	63	101	49			

<sup>a</sup> The scores represent the closest approximation into quartiles that was possible except for Factor 4 for women and Factor 5 for men where only thirds were possible. The abbreviations H, M, L stand for high, medium, and low, respectively.

<sup>b</sup> Categorization based on highest educational level begun.

<sup>c</sup> N's vary due to incomplete data on items comprising the several factor scores.

several factor scores is clearly confirmed by the data presented in this table. The less educated men would be judged as indicating less distress than the more educated ones if the felt psychological disturbance (Factor 1) and particularly the social inadequacy (Factor 3) scores were considered. But, they would be judged as *more* distressed if the unhappiness (Factor 2), lack of identity (Factor 3), or physical distress (Factor 5) scores were considered. It is interesting to note that some of these findings—specifically those relating to psychological disturbance, social inadequacy, and physical distress—support much of the thinking that has been done with respect to social class differences in the expression of psychological distress. Using education as an index of social status,

the findings conform to hypotheses that have been offered about psychological symptoms and intrapunitive being more common to the middle class, and bodily symptoms being a more characteristic working class mode of expression.

Women with differing levels of education also respond differentially to the several categories of self-evaluations of distress represented by the factor scores. The less educated women indicate *more* felt psychological disturbance (Factor 1) and unhappiness (Factor 2), but *less* social inadequacy (Factor 3) than the more educated women. Education is unrelated to Factor 4—lack of identity—for women.

As can be seen, these relationships differ in some ways from those for men. Particu-

larly, they point up a difference between the symptom factors for men and women. The computation of factor scores in a manner that accentuated the uniqueness of each factor had the effect of emphasizing the different determinants of the responses to symptom checklists for men and women. For men, there appear to be somewhat distinct psychological and physical components involved in the responses, as reflected both in the emergence of Factor 5 and in the differential relationships of Factors 1 and 5 with education; for women, only one symptom factor emerged. The use of a symptom checklist with women, like the summary symptom score which is the basis for their Factor 1 scores, clearly leads to the conclusion that the less educated groups are more distressed. For men this negative relationship is obtained only for the Factor 5 scores, where only the physical components of the symptom checklist (physical ill health) are included in the factor score. When the factor score is based on the more psychological components of a symptom checklist (Factor 1), a clear relationship with education is not obtained; the relationship is slightly positive, with college educated men scoring somewhat higher than grade school educated men. The multiple criterion point, however, remains true for both men and women: the selection of a symptom checklist as the more significant indicator of mental health means that certain groups will be selected as "sick" or "healthy"; a different selection would be made if other criteria for mental health were stressed.

We have highlighted the fact that *different distinctive factors* emerge in this factor analysis of self-evaluations of distress, and have used the analysis of educational subgroup differences to support the need for a multiple criterion approach to subjective adjustment. It is still possible, however, that one might find one factor or perhaps a pattern of the separate factors of self-evaluation which, through a method of external validation, might be chosen as *the* best indicator of mental illness or health. But the problem in such an attempt lies in determining what the external criterion should be. It is difficult to choose such an external validation criterion given the considerable conceptual disagreement as to what the definitions and criteria of

mental illness and health should be. Some of these difficulties are illustrated in the attempt to single out symptom lists as the diagnostic tool, with psychiatrists' judgments being the external validating criterion. As we have seen, this seems to minimize certain aspects of psychic dysfunctioning which many would agree should be included in any psychiatric judgment of healthy or neurotic functioning.

Although the problem is partly a function of the imperfections of our knowledge in this area, it is likely that the many different relevant criteria of disturbance can never be encompassed within a single all-embracing criterion. In attempting to predict to some external criterion from the kinds of factors differentiated in this study, it is probably more appropriate to think of different factors being relevant for different predictions than to think in terms of a single criterion and prediction. For example, if we wish to predict the type of disturbance that seems to be implicit in psychiatrists' diagnoses, subjective evaluations that involve a heavy component of psychophysical symptomatology would be the best predictors. On the other hand, for predicting the kind of impairment that leads to immobilization, breakdown, and the self-diagnosis involved in going for help with one's problems we would look to a quite different set of subjective adjustment measures—for example, those heavily loaded on the inadequacy factor. And for a prediction of "positive" mental health still other factors might be most relevant.

In comparing the factor structures of the men and women in our study, both the similarities in the structures and the differences can have important implications. The communality between the two factor structures points to the important psychological dimensions of subjective distress that can transcend social roles and cultural subgroups. If one were looking for the basic underlying psychological parameters of distress, this focus on the communality would be of greatest importance. A logical step in such a focus would involve factor analyses on specific social groups—different age groupings, different educational statuses, different income levels—to pinpoint even further those communalities that exist in all groups.

What are the implications of finding dif-

ferences in factors for men and women? Although both sexes may encounter similar problems in growing up and in current living, they undoubtedly have learned to deal with these problems in different ways. Differences in factors for men and women may reflect critical ways in which the sexes differ in handling or expressing psychic difficulties. For example, since the physical ill health symptom index appears as part of the felt psychological disturbance factor for women but is separate for men, one perhaps has support for the view that the externalization of psychic disturbance into somatic distress is more common for men than for women. Analyses which focus on such differences in factors can pave the way to a better delineation of the critical differential experiences of men and women, or of any other groups so compared.

#### SUMMARY

A representative sample of the resident population of adults in the United States was interviewed about their feelings of distress in different life areas: perceptions of the self; symptoms of distress; adjustment in marriage, parenthood, and work. Nineteen indices were constructed for special subpopulations—married women with children and employed married men with children. In addition, three indices related to aspects of work adjustment were used only for these men. The indices were intercorrelated and the intercorrelations were factor analyzed separately for men and women.

Five distinctive factors emerged for men and four for women. There was considerable apparent overlap between the two factor structures. For both men and women the factors were identified as: felt psychological disturbance (Factor 1); unhappiness (Factor 2); social inadequacy (Factor 3); lack of identity (Factor 4). For men the fifth factor was labeled physical distress.

The emergence of several distinct factors underlying self-evaluations of distress was discussed as reflecting the need for a multiple criterion approach to the definition of mental health. In particular it was noted that findings about differential distress in population subgroups depend upon the particular syndrome of distress that is used as the criterion

of mental health; this point was illustrated for educational subgroupings.

The overlapping factor structures of men and women point to basic dimensions of distress applicable to all people. The differences in the factor structures for men and women point to major sex differences in the experience of distress. In particular, physical distress seems to be a distinctive factor for men while it is tied to psychological symptoms for women.

#### REFERENCES

- EATON, J. W. The assessment of mental health. *Amer. J. Psychiat.*, 1951, 108, 81-90.
- FIEDLER, F. E., DODGE, JOAN S., JONES, R. E., & HUTCHINS, E. B. Interrelations among measures of personality adjustment in nonclinical populations. *J. abnorm. soc. Psychol.*, 1958, 56, 345-351.
- GURIN, G., VEROFF, J., & FELD, SHEILA. *Americans view their mental health*. New York: Basic Books, 1960.
- HÉRZBERG, F., MAUSNER, B., & SNYDERMAN, BARBARA. *The motivation to work*. New York: Wiley, 1959.
- INKELES, A. Industrial man: The relation of status to experience, perception, and value. *Amer. J. Sociol.*, 1960, 66, 1-31.
- JAHOA, MARIE. Toward a social psychology of mental health. In J. M. E. Senn (Ed.), *Symposium on the healthy personality*. New York: Josiah Macy, Jr. Foundation, 1950. Pp. 211-230.
- JAHOA, MARIE. The meaning of psychological health. *Soc. Casewk.*, 1953, 34, 349-354.
- JAHOA, MARIE. *Current concepts of positive mental health*. New York: Basic Books, 1958.
- KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
- KISH, L. Selection of the sample. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953. Pp. 175-240.
- MACMILLAN, A. M. The health opinion survey: Technique for estimating prevalence of psychoneurotic and related types of disorder in communities. *Psychol. Rep.*, 1957, 3, 325-329.
- MORSE, NANCY C. *Satisfactions in the white-collar job*. Ann Arbor: Survey Research Center, 1953.
- ROGERS, C. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
- SCOTT, W. A. Research definitions of mental health and mental illness. *Psychol. Bull.*, 1958, 55, 67-87.
- SMITH, M. B. Optima of mental health: A general frame of reference. *Psychiatry*, 1950, 13, 503-510.
- SMITH, M. B. Research strategies toward a conception of positive mental health. *Amer. Psychologist*, 1959, 14, 673-681.
- TERMAN, L. M., & WALLIN, P. The validity of marriage prediction and marital adjustment tests. *Amer. sociol. Rev.*, 1949, 14, 497-505.

(Received February 8, 1961)



## DIFFERENCES AND DISAGREEMENT AS FACTORS IN CREATIVE GROUP PROBLEM SOLVING<sup>1</sup>

L. RICHARD HOFFMAN, ERNEST HARBURG, AND NORMAN R. F. MAIER

*University of Michigan*

Differences in opinions and ideas among group members need not lead to dissatisfactions and unpleasant experiences but rather can lead to invention and creative problem solving (Maier, 1958). When opposing viewpoints are brought together in a free exchange, the resulting conflict may be resolved by the creation of a new solution which either reorganizes or incorporates the different viewpoints. This solution becomes a new meeting ground and is therefore satisfactory to all. Recent studies have provided support for these propositions by showing that differences among group members promote effective problem solving (e.g., Ghiselli & Lodahl, 1958; Hoffman, 1959; Hoffman & Maier, 1961a).

The mere existence of different viewpoints in a group does not guarantee their full expression or consideration, however, especially if one of the viewpoints is held by a person of authority over the other group members. An authority figure is likely to dominate the discussion and force his ideas on the group, disregarding or even suppressing the opposing ideas of his subordinates (Maier & Hoffman, 1960, 1961; Torrance, 1955). His dominance interferes with the free expression of differing opinions and reduces the possibility of the emergence of creative solutions to problems.

The usual approach to overcome the dominance of a person in authority over discussions is to attempt to reduce his influence in some way. Maier (1952) has suggested that group leaders refrain from suggesting ideas, provide expert information as needed, and allow subordinates to make all suggestions for problem solutions. In this way the leader's favored idea is less likely to dominate the discussion than if he makes his views known be-

fore his subordinates have had their chance to be heard.

A less common way of ensuring the expression of differing opinions by subordinate members of the group is to increase their commitment to those opinions. Subordinates are likely to give strong expression to their viewpoints when they feel sufficiently secure to disagree with the person in authority, as when they belong to a strong labor union, or when they have sufficiently strong feelings about the subject being discussed. Increasing the subordinates' commitment so that their point of view can effectively oppose that of the authority figure, thus may be regarded as a way to introduce conflict and to provide the conditions for creative problem solving.

These two approaches to ensuring the free expression of different opinions in the group—increasing the subordinates' commitment and reducing the superior's dominance—were combined in the present experiment to determine their separate and combined effectiveness in improving the quality of group solutions to a problem.

The Change of Work Procedure problem (Maier, 1952)—a role playing situation involving a conflict between a foreman and his subordinate workers about the relative merits of two work methods—was selected for use in this experiment, since it seemed to provide the conditions needed to test the experimental variations. Experience with this problem in its standard version has shown that the foreman's preferred work method is adopted in a majority of the cases and creative solutions occur relatively infrequently. Previous research also has indicated that the production of creative solutions to the problem was increased when subjects' orientation to authority relations was low (Maier & Hoffman, 1961), when foremen were trained to be permissive (Maier, 1953), and when the groups solved the problem a second time (Maier &

<sup>1</sup> This investigation was supported by a USPHS research grant (M-2704) from the National Institute of Mental Health, United States Public Health Service.

Hoffman, 1960). The improvement in the last study seemed to result principally from the foreman's greater willingness to permit the workers to influence the second decision, since a solution had been reached in the first meeting.

Variations in the relative dominance of the foreman and in the strength of the workers' feelings were introduced in the present study by altering the degree of motivational support for each side's position, as described in the role instructions given to each group member. The workers were provided with reasons to be either more or less committed to the work method they were currently using. The foremen were given the basis either for strongly encouraging the adoption of a new work method or for being more considerate of the workers' feelings in suggesting its adoption.

If conflict is potentially conducive to effective problem solving, high quality solutions should be produced more often by groups where the workers are more committed to the existing work method than where they are less committed, since conflict is more likely to occur in the former condition than in the latter. Conflict can also lead to personal animus instead of invention, especially if the foreman does not see conflict as offering the possibility of invention. High quality solutions are thus more likely to appear where the foreman is considerate in the face of strongly committed workers than where the foreman is himself strongly committed to his own suggestion.

While an initial conflict is hypothesized to set the stage for possible creativity, whether the final result will be a creative solution or the acceptance of one of the less satisfactory solutions will still depend upon the group's ability to discover a creative solution. Even when the foreman is willing to consider alternative solutions, these cannot be adopted unless they are developed by someone. Differences among the group members appear to be important here also, in that they provide a variety of perceptions of the problem and, therefore, different "directions" in thinking (Maier, 1930). A variety of directions is postulated to increase the probability of discovering a creative solution to a problem.

Mixing the sex composition of a group is one way of providing such differences. In a previous study in which the Change of Work Procedure problem was used (Hoffman & Maier, 1961a), mixed-sex groups produced a somewhat (but not significantly) higher proportion of high quality solutions than did all-male groups, despite the fact that all-female groups are usually poor at problem solving. In the present experiment the sex composition of the groups was varied systematically so that the quality of solutions of mixed-sex, all-male, and all-female groups can be compared under each condition of foreman-worker conflict. Mixed-sex groups may be expected to produce more high quality solutions than single-sex groups. Since the availability of different perspectives or directions should be especially valuable where conflict leads to a search for new solutions, the mixed-sex groups may be expected to be especially effective where the workers in the groups are committed to the existing work method, i.e., where conflict occurs.

## METHOD

### *Problem*

The Change of Work Procedure problem (Maier, 1952) is a role playing situation involving a foreman and three of his subordinate workers in an automobile factory. The workers assemble fuel pumps and are paid on a group piece-rate basis. The assembly operation is divided into three simple jobs and the men presently rotate hourly among the three positions. The role playing consists of a meeting called by the foreman to discuss a change in work method in which each worker would work on only one position, his best according to time-study data provided for the foreman in his role instructions.

In a majority of groups, the foreman is able to overcome the workers' fears and resistance and to oversell the group on the virtues of the *new* fixed-position method—principally that the increased production achieved would result in an increase in their pay. In some instances the workers refuse to change and retain the *old* method of working, usually on the grounds that the boredom of working on the new method would cause a decline, or at least no increase, in production. In a still smaller percentage of groups, a third type of work method is produced—a type in which the advantages of the individual differences in worker productivity are exploited, yet boredom is still relieved through rotation. These solutions have been termed *integrative*. Their production represents the release of the group from a situation of choice between the old and new methods to permit a search for additional solutions to

the problem. It is the integrative solutions that are here considered to be evidence of creative problem solving by the group.

### *Establishment of Experimental Conditions*

**Conflict.** The degree of conflict was varied by altering parts of the standard role instructions to the foreman and workers.

All foremen received the following instructions:

You are the foreman in a shop and supervise the work of about 20 men. Most of the jobs are piece-rate jobs and some of the men work in teams and are paid on a team piece-rate basis. In one of the teams, Jack, Walt, and Steve work together. Each one of them does one of the operations for an hour and then they exchange, so that all men perform each of the operations at different times. The men themselves decided to operate that way and you have never given the plan any thought.

Lately, Jim Clark, the methods man, has been around and studied conditions in your shop. He timed Jack, Walt, and Steve on each of the operations and came up with the following facts:

TIME PER OPERATION

	Position 1	Position 2	Position 3	Total
Jack	3 minutes	4 minutes	4½ minutes	11½ minutes
Walt	3½ minutes	3½ minutes	3 minutes	10 minutes
Steve	5 minutes	3½ minutes	4½ minutes	13 minutes
				34½ minutes

He observed that with men rotating, the average time for all three operations would be  $\frac{1}{3}$  of the total time or 11½ minutes per complete unit. If, however, Jack worked in the Position 1 spot, Steve in the Position 2 spot, and Walt in the Position 3 spot, the time would be 9½ minutes, a reduction of over 17%. Such a reduction in time would amount to a saving of more than 80 minutes. In other words, the lost production would be about the same as that which would occur if the men loafed for 80 minutes in an 8-hour day. If the time were used for productive effort, production would be increased more than 20%.

In addition, to encourage commitment to the new work method, half the foremen (referred to hereafter as Dominant foremen) were told in their instructions:

You were very enthusiastic about this idea since it seemed to offer the men an opportunity to make more money without any more effort. At the same time, it will yield higher production for your unit. Although you have been satisfied with their production up to now, you have called Jack, Steve, and Walt in to tell them about this new plan.

You see no reason why they should object to this change in operation.

The rest of the foremen (referred to hereafter as Considerate foremen) were told instead:

This argument made pretty good sense to you so you have decided to take it up with the men. Since their production has been satisfactory up to now, you don't want to make any changes without their approval.

Although these latter instructions were designed to make the foreman favorable to the new method, they also should have made him more considerate of the workers' feelings on the matter.

Parts of the workers' roles were altered also to make them feel more or less bored with the present situation as the conditions required, thus more or less committed to the present work method (the old solution). The workers whose roles made them more committed to the old work method will be called Strong workers and those who were intended to be less committed will be called Weak workers to represent the strength of the feelings involved.

Some examples of the specific changes made in the role instructions will give the reader a sense of their nature. An example of the Weak and Strong instructions is given for each character.

(Weak) Sometimes the work gets to be monotonous.

(Strong) The work is very monotonous.

(Weak) The three of you exchange positions every hour.

(Strong) Because the work is so monotonous, the three of you decided to exchange positions every hour.

(Weak) You have learned to set your pace to music you hum to yourself and the jobs don't bother you at all.

(Strong) You have tried to make them [the jobs] less monotonous by talking, daydreaming, and changing your pace. You have even tried to set your work pace to music you hum to yourself. Each of these seems to help for a short time. . . . the change adds variety to the work.

In addition to these feelings that the workers hold about the monotony of the job, Steve is afraid of losing his job because he is inefficient and Walt is antagonistic toward the time-study man because he feels time study is a waste of the company's money.

Four experimental conditions varying in their degree of conflict were thus developed by combining the two forms of the foreman's role and the two forms of the workers' roles: Dominant foreman-Strong workers; Considerate foreman-Strong workers; Dominant foreman-Weak workers; Considerate foreman-Weak workers.

**Sex Composition.** The sex composition of the groups was varied systematically within each conflict condition by assigning eight groups each of all-male, all-female, and mixed-sex (two males and two females) randomly to each conflict condition. Thus



the 96 four-person groups were assigned to the following conditions with eight groups in each cell:

Foreman:	Dominant		Considerate	
Workers:	Strong	Weak	Strong	Weak
Sex Composition:	Male	Male	Male	Male
	Mixed	Mixed	Mixed	Mixed
	Female	Female	Female	Female

Also, in each conflict condition the foreman's role was assigned to males in half the mixed-sex groups and to females in the other half.

*Other Measures.* Subjects were asked about their satisfaction with the solution reached by their group and with their amount of influence over the solution, using Likert-type six-point questions. Values from 1 to 6, indicating increasing degrees of satisfaction were assigned to the alternatives: very dissatisfied, somewhat dissatisfied, neither satisfied nor dissatisfied, fairly satisfied, quite satisfied, very satisfied.

### Subjects

Three hundred and eighty-four subjects, half of each sex, were recruited from various freshman and sophomore level courses at the University of Michigan. The problem was multiple role played (Maier & Zerfoss, 1952) in 10 sessions.

## RESULTS

### Conflict Conditions

The principal hypothesis of this study is that an increase in the strength of the workers' feelings about the monotony of their jobs causes them to oppose effectively the foreman's suggested change and that the resulting conflict generates a higher proportion of integrative solutions. Table 1 presents the proportions of solutions of each type produced by groups in each conflict condition, as well as separately by the type of foreman and type of workers. A comparison of the results for groups with Strong as against Weak workers supports the principal hypothesis at the .01 level of confidence by *t* test. Integrative solutions were produced by 45.8% of the groups with Strong worker roles and by only 18.8% of the groups with Weak worker roles, disregarding the type of foreman's role. The resistance of the Strong workers to the foreman's favored solution (new) is apparent from the table. Less than half as many new solutions were accepted by groups with Strong workers as by groups with Weak workers. This suggests that integrative solutions re-

sulted when the point of view opposing the foreman's was strongly held and a conflict developed which forced the groups to create different types of solutions as a way of reaching agreement.

On the other hand, Table 1 also shows that altering the foreman's role had little effect. Only a slightly higher proportion of integrative solutions was produced by groups with Considerate foreman's roles (35.4%) as compared with groups with Dominant foreman's roles (29.2%), disregarding the type of workers' roles. Groups accepted new solutions more frequently than either old or integrative solutions under both types of foremen, although the Dominant foremen were somewhat more successful in having their way.

When the workers had weak feelings about the monotony of the jobs, even the Considerate foremen seem to have dominated the discussion, as evidenced by the fact that the same high proportion of new solutions (66.7%) was produced by the Dominant and Considerate foremen groups in the Weak workers condition. Where the workers' feelings were strong, on the other hand, the Considerate foremen appear to have been somewhat more responsive to them than the Dominant foremen were (Table 1).

TABLE 1  
DISTRIBUTIONS OF SOLUTIONS BY GROUPS IN EACH CONFLICT CONDITION

Type of solution	Type of group			
	Strong workers (N=48)	Weak workers (N=48)	Dominant foreman (N=48)	Considerate foreman (N=48)
Integratives New Old	45.8% 31.2 22.9	18.8% 66.7 14.6	29.2% 54.2 16.7	35.4% 43.8 20.8
	99.9%	100.1%	100.1%	100.0%
Foreman: Workers:	Dominant Strong (N=24)	Considerate Strong (N=24)	Dominant Weak (N=24)	Considerate Weak (N=24)
	37.5% 41.7 20.8	54.2% 20.8 25.0	20.8% 66.7 12.5	16.7% 66.7 16.7
Integratives, <sup>a</sup> New <sup>b</sup> Old <sup>b</sup>	100.0%	100.0%	100.0%	100.1%

<sup>a</sup> Tests of differences in percentage of integrative solutions produced by groups with: Strong vs. Weak workers—*t* = 2.93, *p* < .01; Dominant vs. Considerate foremen—*t* = 0.65, *ns*; Strong workers with Dominant foreman vs. Considerate foremen—*t* = 1.15, *ns*.  
<sup>b</sup> For relationship among conflict conditions and type of solution: chi square = 11.30, .10 > *p* > .05.

TABLE 2  
SOLUTIONS IN GROUPS OF DIFFERENT  
SEX COMPOSITION

Type of solution	Type of group							
	Mixed sex		Single sex <sup>a</sup>		All male		All female	
	N	%	N	%	N	%	N	%
Integrative <sup>b</sup>	13	40.6	18	28.1	12	37.5	6	18.8
New	10	31.2	37	57.8	17	53.1	20	62.5
Old	9	28.1	9	14.1	3	9.4	6	18.8

<sup>a</sup> Single sex combines the results of the all-male and all-female groups.  
<sup>b</sup> Test of difference in percentage of integrative solutions produced by mixed-sex and single-sex groups,  $z = 1.19$ ,  $ns$ .

Sex Composition

Table 2 shows the distributions of solutions produced by groups of each sex composition, regardless of the conflict conditions. While a  $z$  test of the difference in the proportions of integrative solutions produced by mixed-sex groups and single-sex groups is not significant ( $z = 1.19$ ,  $p = .15$ ), certain differences are apparent. The mixed-sex groups produced the lowest proportion of new solutions, which raised their proportions of both integrative and old solutions.<sup>2</sup> This result parallels in more modest fashion the effect produced by strengthening the workers' commitment to the old method (see Table 1). Also, the lowest proportion of integrative solutions was produced by the all-female groups.

Interaction of Conflict Conditions and Sex Composition

An examination of the combined effects of the conflict conditions and the sex composition of the groups indicates that the trend shown in Table 2 was even more marked in the strong worker condition. The mixed-sex groups in the Strong worker condition produced the smallest proportion of new solutions and these were distributed between integrative and old solutions (Table 3). However, an insufficient number of the mixed-sex, Strong worker groups produced integrative solutions, as compared with the number pro-

<sup>2</sup> The distributions of solutions in mixed-sex groups led by male foremen were not different from those led by female foremen, so the results will be reported together.

duced by single-sex, Strong worker groups, to confirm statistically the hypothesis that these two variables would interact to produce a substantially higher proportion of creative solutions. The proportions of integrative solutions produced by these two types of groups were not statistically significant ( $z = 1.15$ ,  $p = .15$ ).

A chi square test of the over-all relationship among sex composition, worker commitment, and type of solution is statistically significant (chi square = 15.58,  $p < .05$ ). The significance arises from the small proportion of new solutions (6.2%) accepted by the mixed-sex, Strong worker groups, with an accompanying increase in the production of integrative and old solutions, as compared to the high proportion of new solutions (71.9%) obtained from the single-sex, Weak worker groups. Moreover, the linear increase in the proportion of integrative solutions produced as the conflict and sex composition factors are added offers suggestive support for the hypothesis concerning the interaction of these variables. The differences in proportions of integrative solutions produced by the Weak worker, single-sex groups and both the Strong worker, mixed-sex and Strong worker, single-sex groups are significant at the .01 level of confidence, while the differences between these groups and the Weak worker, mixed-sex groups were not significant ( $z$  tests).

One other point of interest in Table 3 is that the six integrative solutions produced by the all-female groups all occurred in the Strong worker conditions.

Satisfaction with Influence and with the Solution

The basic assumption underlying this experiment is that the free interplay of the different ideas present in a group is fundamental to the development of high quality solutions to problems. The attempt to increase the workers' commitment to their present work method was designed to improve the quality of the solutions by permitting the workers' views to be placed in effective opposition to the foreman's, and, thus force their consideration by the group. The advantages and disadvantages of both the foreman's and the workers' viewpoints can thereby be discov-

TABLE 3  
SOLUTIONS BY STRONG AND WEAK WORKER GROUPS OF DIFFERENT SEX COMPOSITION

SOLUTIONS BY STRONG AND WEAK WORKER GROUPS OF DIFFERENT TYPE																
Type of solution <sup>a</sup>	Type of group															
	Strong workers								Weak workers							
	Mixed sex		Single sex		All male		All female		Mixed sex		Single sex		All male		All female	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Integrative <sup>b</sup>	9	56.2	13	40.6	(7)	(43.8)	(6)	(37.5)	4	25.0	5	15.6	(5)	(31.2)	(0)	(0.0)
New	1	6.2	14	43.8	(6)	(37.5)	(8)	(50.0)	9	56.2	23	71.9	(11)	(68.8)	(12)	(75.0)
Old	6	37.5	5	15.6	(3)	(18.8)	(2)	(12.5)	3	18.8	4	12.5	(0)	(0.0)	(4)	(25.0)

<sup>a</sup> Integrated vs. single), and type of solution: chi square = 15.58.

<sup>a</sup> For relationship among worker commitment, sex composition (mixed vs. single), and type of solution: chi square = 15.58,  $p < .05$ . The numbers in parentheses were not included in this calculation.

<sup>b</sup> Test of difference in percentage of integrative solutions produced by Strong worker groups of mixed- vs. single-sex.  $z = 1.00$ ,  $ns$ .

ered and a joint effort can be made to create a new solution, which will incorporate the advantages and minimize the disadvantages. One measure of the extent to which the problem solving process was a joint one, rather than unilateral, would be that both the foreman and the workers were satisfied with the amount of influence they had over the solution.

Groups were classified as being mutually satisfied with their influence over the solution, according to the subjects' responses to the question about their satisfaction with influence, where the foreman and all the workers checked responses 5, quite satisfied, or 6, very satisfied. According to this stringent criterion, 12 of the 96 groups were mutually satisfied with their influence over the solution. Table 4 shows that 8 of these 12 groups (66.7%) produced integrative solutions, compared to only 23 of the 84 groups not mutually satisfied with their influence (27.7%). This difference is significant at the .05 level of confidence. Interestingly, 6 of the 12 groups (50.0%) mutually satisfied with their influence over the solution occurred in the Considerate foreman-Strong worker condition, and all 6 produced integrative solutions. It is apparent that the Considerate foremen did not capitulate to the Strong workers' feelings and often successfully initiated change through a process of joint problem solving.

The method of increasing worker influence used in this experiment did not necessarily ensure acceptance of the integrative solutions thereby produced, however. Table 5 shows

the percentage of groups in which the foremen and workers, respectively, were satisfied with each type of solution and in which they were mutually satisfied with the solution. It will be seen that the foremen were least often satisfied with the old solutions ( $p < .05$ ,  $z$  test) and the workers were least often satisfied with the new solutions ( $p < .05$ ,  $z$  test). The proportions of foremen and workers satisfied with the integrative solutions were no greater than the proportions satisfied with the solution they preferred initially. However, the same proportion of foremen was satisfied with the integrative solutions as with the new ones, and only a somewhat higher proportion of workers was satisfied with the old solutions than with the integrative.

TABLE 4  
SOLUTIONS AND SATISFACTION WITH INFLUENCE OVER THE SOLUTION

Type of group	Type of solution					
	Old		New		Integrative <sup>a</sup>	
	N	%	N	%	N	%
Mutually satisfied with influence over solution <sup>a</sup>	1	8.3	3	25.0	8	66.7
Not mutually satisfied with influence over solution	16	19.3	44	53.0	23	27.7

<sup>a</sup> Groups were classified as mutually satisfied with influence when the foreman and all the workers responded 5, quite satisfied, or 6, very satisfied. One group could not be classified because its foreman did not indicate his satisfaction with influence over an old solution.

\* Percentage difference significant at the .05 level of confidence,  $z = 2.11$ .



TABLE 5  
TYPE OF SOLUTION AND SATISFACTION  
WITH THE SOLUTION

	Type of solution					
	Old		New		Integrative	
	N	%	N	%	N	%
Foremen <sup>a</sup>						
Satisfied <sup>b</sup>	3	17.6*	30	63.8	20	64.5
Not satisfied	14	82.4	17	36.2	11	35.5
Workers						
Satisfied <sup>b</sup>	41	75.9	70	49.6*	55	59.1
Not satisfied	13	24.1	71	50.4	38	40.9
Satisfaction in groups <sup>a</sup>						
Mutually satisfied <sup>b</sup>	2	11.8	4	8.5	5	16.1
Not mutually satisfied	15	88.2	43	91.5	26	83.9

<sup>a</sup> One foreman did not indicate his satisfaction with an old solution.

<sup>b</sup> Responses of 5, quite satisfied, and 6, very satisfied, were classified as satisfied; all others as not satisfied. Groups were classified as mutually satisfied where the foreman and all the workers were satisfied.

\* Percentage satisfied is smaller for this solution than for both others at the .01 level of confidence.

The foremen and workers were somewhat more likely to be mutually satisfied with the integrative solutions than with either the old or new solutions. Also, four of the five groups mutually satisfied with their integrative solutions were also mutually satisfied with their influence over the solution, while only one of the six groups mutually satisfied with the old or new solutions were also mutually satisfied with their influence. These results suggest that where the foreman and workers freely exchanged ideas in solving the problem, both the quality and acceptance of the solutions were improved.

## DISCUSSION

This experiment was designed to test the hypothesis that conflict of ideas causes groups to search for alternatives and thereby to improve the quality of their solutions. The results support this hypothesis. Increasing the workers' resistance in this situation, where the foreman usually dominates the decision, resulted in an increased proportion of integrative solutions. On the other hand, when the workers did not feel strongly about the monotony of their jobs, they submitted fairly readily to the foreman's request to adopt the

new work method and the groups rarely developed integrative solutions.

The data suggest, however, that if high acceptance as well as high quality of the solutions is to be achieved, the conflict of ideas must occur within an atmosphere of tolerance for others' points of view. Where such an atmosphere was lacking, the conflict of ideas which occurred when the workers held strong feelings often produced integrative solutions which lacked strong member satisfaction with the outcome. The struggle involved in arriving at a solution may have caused hard feelings in the group and have lowered the group's enthusiasm for the agreed-upon work method.

The lack of strong over-all effect of making the foreman Considerate rather than Dominant, especially where the workers' feelings were Weak, reinforces the importance of conflict for achieving high quality solutions. The Weak workers offered little opposition to the foreman's proposition even where he was willing, supposedly, to listen to their views. Without strong opposition to the new method, there was no cause for the group to consider possible alternative solutions. The foreman was able to convince the workers of the benefits of the new method, and by virtue of his authority over them, get them to accept it on a trial basis even though they were not satisfied with the method. To achieve creative solutions to problems under these conditions, a superior should probably encourage the expression and consideration of opposing points of view, rather than try to persuade the group of the merits of his own ideas. There was some suggestion in the data from groups with Strong workers that the Considerate foremen were more likely to achieve integrative solutions than were the Dominant foremen, possibly because they were more prone to consider seriously the opposition point of view.

The interactive effects on problem solving of the sex composition of the groups and the conflict conditions deserve comment. Since women are generally poorer problem solvers than men, one might have expected them to reduce the effectiveness of the males in the mixed-sex groups. Thus the proportion of integrative solutions from the mixed-sex groups

might have been expected to have been approximately the mean proportion produced by the all-male and all-female groups. Instead, the proportion of integrative solutions produced by the mixed-sex groups was slightly greater than that produced by the all-male groups, especially in the Strong worker condition, and substantially greater than the proportion produced by the all-female groups.

These results provide further suggestive evidence to support the proposition that part of women's usual poorer problem solving performance is due to a lack of motivation related to their sex-role identification (Milton, 1957, 1959). Hoffman and Maier (1961b) had shown previously that mixed-sex groups seemed to stimulate their female members to greater and more effective effort. In the present study, both the slightly improved performance of the mixed-sex groups in the Strong worker condition and the fact that all the integrative solutions produced by all-female groups were produced in the Strong worker condition, suggest that these role instructions may have produced a higher degree of involvement in the problem among the female subjects.

The general findings of this study seem to support the following conclusions:

1. Different perspectives lead to different "directions" for identifying a problem (Maier, 1930) and to alternative possibilities for solutions and, therefore, offer the potential for creative problem solving.
2. Different perspectives, when set in effective opposition to each other, can delay the premature acceptance of an obvious solution and contribute to turning a choice situation into a problem. A group may then be forced to search for alternative solutions which might better satisfy the requirements of the problem.
3. The joint resolution of such conflict by all members of the group leads to solutions of high quality which are well accepted by the group members.

#### SUMMARY

This experiment was designed to test the following hypothesis: conflict generated by increasing the resistance of subordinate mem-

bers of a group to a solution offered by a member in an authority position increases the frequency of high quality solutions. The supplementary hypothesis that mixed-sex composition in interaction with such conflict further enhances the production of high quality solutions was also tested.

Ninety-six four-person groups role played the Change of Work Procedure problem under conditions where a foreman was either dominant or considerate in combination with subordinate workers whose commitment to the *status quo* was either relatively strong or weak. Eight groups each of all-male, all-female, and mixed-sex composition were assigned at random to each condition.

The results support the major hypothesis. High quality solutions were produced by 45.8% of groups under Strong worker conditions and by only 18.8% of groups under Weak worker conditions. The hypotheses concerning the effects of sex composition were only supported suggestively.

The discussion emphasizes the importance for achieving problem solutions of both high quality and acceptance, of encouraging the expression and consideration of conflicting points of view in an atmosphere of mutual influence among the group members. Differences of opinion can promote creative problem solving by turning a choice situation into a problem and by providing a variety of directions in thinking.

#### REFERENCES

- GHISELLI, E. E., & LODAHL, T. M. Patterns of managerial traits and group effectiveness. *J. abnorm. soc. Psychol.*, 1958, 57, 61-66.
- HOFFMAN, L. R. Homogeneity of personality and its effect on group problem-solving. *J. abnorm. soc. Psychol.*, 1959, 58, 27-32.
- HOFFMAN, L. R., & MAIER, N. R. F. Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *J. abnorm. soc. Psychol.*, 1961, 62, 401-407. (a)
- HOFFMAN, L. R., & MAIER, N. R. F. Sex differences, sex composition, and group problem solving. *J. abnorm. soc. Psychol.*, 1961, 63, 453-456. (b)
- MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, 10, 115-143.
- MAIER, N. R. F. *Principles of human relations*. New York: Wiley, 1952.
- MAIER, N. R. F. An experimental test of the effect of training on discussion leadership. *Hum. Relat.*, 1953, 6, 161-173.

- MAIER, N. R. F. *The appraisal interview: Objectives, methods and skills*. New York: Wiley, 1958.
- MAIER, N. R. F., & HOFFMAN, L. R. Quality of first and second solutions in group problem solving. *J. appl. Psychol.*, 1960, **44**, 278-283.
- MAIER, N. R. F., & HOFFMAN, L. R. Organization and creative problem solving. *J. appl. Psychol.*, 1961, **45**, 277-280.
- MAIER, N. R. F., & ZERFOSS, L. F. MRP: A technique for training large groups of supervisors and its potential use in social research. *Hum. Relat.*, 1952, **5**, 177-186.
- MILTON, G. A. The effects of sex-role identification upon problem solving skill. *J. abnorm. soc. Psychol.*, 1957, **55**, 219-244.
- MILTON, G. A. Sex differences in problem solving as a function of role appropriateness of the problem content. *Psychol. Rep.*, 1959, **5**, 705-708.
- TORRANCE, E. P. Some consequences of power differences on decision making in permanent and temporary three-man groups. In A. P. Hare, E. F. Borgatta, & R. F. Bales (Eds.), *Small groups*. New York: Knopf, 1955. Pp. 482-492.

(Received February 14, 1961)



## PERCEPTUAL RECOGNITION AS A FUNCTION OF ANXIETY, MOTIVATION, AND THE TESTING SITUATION<sup>1</sup>

E. D. LONGENECKER<sup>2</sup>

Pennsylvania State University

Common to most studies investigating the influence of anxiety on behavior is the fundamental hypothesis that anxiety acts primarily as a drive stimulus. In the work of Spence and Taylor and their associates involving simple, one response situations such as conditioning there appears to be a direct relationship between anxiety and performance level. Their assumption is that anxiety level reflects a difference in the subject's drive which combines with the habit strength of the dominant response in the subject's response hierarchy to raise and strengthen the excitatory potential. On the other hand, in more complex problem solving situations where no single response dominates the response hierarchy, anxiety level is found to be inversely related to performance. The assumption here is that the higher anxiety level increases the response strength for all responses in the response hierarchy. Since the strength of competing responses will be raised accordingly, the probability of the occurrence of the correct response will be lowered. Put another way, the more competing responses brought to a threshold level, as one's drive level is increased, the more interference there will be with the completion of the problem solving task.

Mandler and Sarason (1952) approach the problem somewhat differently. They postulate that two fundamental drives are activated simultaneously in any problem solving situation. First, there is an anxiety drive which is capable of eliciting all responses in the individual's response hierarchy regardless of their relevance to the completion of the

task. It is assumed that the task relevant responses are specific to each task and that their habit strength is built up only as a result of the reinforcement gained upon successful completion of each item on the task. In addition, there are the task irrelevant responses which are assumed to be competing responses existing in the individual's hierarchy of responses as a result of past experiences. These *competing, task irrelevant responses* are generalized to each new task and are assumed to hinder successful performance.

Thus it is postulated that, at the beginning of a serially presented task when the strength of the relevant responses is low, anxiety functions as a negative drive. As the task relevant responses are reinforced and the generalized irrelevant responses are not reinforced, however, anxiety gradually comes to function as a positive drive. In a series of timed, perceptual recognition items it would be predicted that high anxious (HA) subjects would have significantly slower reaction times on the early trials than would low anxious (LA) subjects. As the succeeding items of the tests are successfully completed, however, the reaction times of the HA subjects would gradually approach and eventually become significantly faster than those of the LA subjects.

A second drive which Mandler and Sarason (1952) believe to be operating in a testing situation they call the "task drive," perhaps akin to achievement motivation. According to these authors, this drive elicits only the so-called task relevant responses, or those responses leading to successful completion of the task. Thus a direct relationship between the strength of the task drive and performance on a given task is postulated. Consequently, in our recognition tests it would be predicted that subjects with a high task drive (i.e., high motivation, HM), would have significantly faster reaction times than would

<sup>1</sup> This study is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Texas, 1956.

<sup>2</sup> The author wishes to express his sincere appreciation for the encouragement and direction given by his committee chairman, Wayne H. Holtzman, throughout the study.

subjects with a low task drive (i.e., low motivation, LM).<sup>3</sup>

Although Mandler and Sarason (1952) state that anxiety and motivation (i.e., their task drive) are influenced by the nature of the testing situation, instructions, etc., they do not state precisely how these variables are expected to influence these drives. On the assumption that thresholds for the arousal of both anxiety and motivation are lower for subjects who are high on these variables than they are for subjects who are low on these variables, it may be expected that ego involving conditions would have a differential effect on both the subjects high or low in anxiety and the subjects high or low in motivation. To illustrate: stress conditions will heighten the functional level of anxiety in the HA subjects but will not heighten the functional level of the anxiety drive in the LA subjects. Thus it would be predicted that HA subjects given a series of tests under stress conditions would have significantly slower reaction times on the early trials of each test than HA subjects given the tests under nonstress conditions. On subsequent trials of each test, this significant difference would eventually reverse itself. In contrast, LA subjects given the tests under stress conditions would have reaction times that are not significantly different from LA subjects given the tests under nonstress conditions.

Applying the same assumption to the variable of motivation, stress conditions will heighten the level of motivation of the HM subjects but will not heighten the level of motivation of the LM subjects. Thus it would be predicted that HM subjects given the tests under stress conditions would have reaction times that are significantly faster than for the HM subjects given the tests under nonstress conditions. However, the LM subjects given the tests under stress conditions would have reaction times that would not be significantly different from LM subjects given the tests under nonstress conditions.

The following study is designed to investigate these conceptualizations of the functions of anxiety, a task oriented drive (i.e., motivation), and stress.

<sup>3</sup> For simplicity's sake, only the term Motivation will be used in referring to this variable throughout this paper.

## METHOD

### *Experimental Design and Subjects*

The basic design is a  $3 \times 2$  analysis of variance with the variables of Anxiety and Motivation being dichotomized into high and low groups while the third variable represents Stress and Nonstress testing conditions. All subjects were sophomore male students from the University of Texas with ACE scores within the range from 90 to 120 inclusive (approximately the middle 50% of the students admitted to the university) who met the criteria defining the variables of Anxiety and Motivation. The experimental control of sex and intelligence was necessitated by the significant correlations reported by several authors (Mandler & Cowen, 1958; Rockett, 1955; Sarason, 1957a, 1957b, 1959; Sarason & Mandler, 1952; Thurstone, 1944; Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954) between these variables and the criterion measures and the perceptual recognition tests used in this study.

### *Procedure*

*Measures of anxiety and motivation.* HA and LA subjects were selected on the basis of their extreme scores on the Sarason Scale of Test Anxiety. The cutoff scores used to select these two groups represented the 28th and the 70th percentile points on the distribution of scores for a randomly selected group of 292 male subjects with ACE scores as defined above. The main reason for preferring this measure of anxiety over the other measures now available is based on the fact that any measure of "anxiety" only measures what might be called "anxiety proneness" to certain stimulus situations. Thus it would seem important that the specific stimulus situation used in the experiment to arouse anxiety be closely related to the criterion measure.

There were several other reasons for using this scale. First, this scale has been standardized on a college population, and Gordon and Sarason (1955) report split-half and test-retest reliabilities of .91 and .82, respectively. Furthermore, its disguise as a measure of test anxiety (it is administered as an attitude questionnaire) and its ability to elicit the interest of the subjects completing it seemed most desirable.

HM and LM subjects were selected on the basis of a combination of under- and overachievement in college grades and low and high need Achievement scores on the Edwards Personal Preference Schedule (PPS). In selecting subjects with records of either under- or overachievement, students having either a grade point ratio (GPR) of 1.1 or less (i.e., underachievers) or a GPR of 1.9 or greater (i.e., overachievers) were selected with GPR being measured on a three-point scale. On the need Achievement scale of the Edwards PPS, the second measure used in defining motivation, cutoff scores of 13 and 18 were selected. These cutoff scores represented the 35th and the 72nd percentile points, respectively, on the distribution of scores for the 292 subjects completing the criterion



measures. Thus for a subject to be included in the LM group, he had to have a low GPR and a low score on the need Achievement scale; and for a subject to be included in the HM group he had to have a high GPR and a high score on the need Achievement scale.

The reason for using this dual criterion measure for motivation stems primarily from the distinction between the concepts of "need" and "drive." It is conceivable that certain subjects will have a high "need" for achievement but will be unable, for some reason or other, to exert themselves over a long period of time such as is demanded in maintaining a GPR up to their levels of capacity. For the purposes of the present study, it seemed important to obtain students who not only indicated a high or low level of aspiration for achievement as is most likely to be shown by the Edwards PPS score but also to have some evidence that they have shown a "drive" level which has been consistent with the "need" level. Thus it seemed wise to select only those subjects scoring either high-high or low-low in terms of their "expressed need" to achieve and their actual achievement.

*Relationships of criterion measures.* Since four different measures were obtained in defining the criterion variable of Anxiety and Motivation, it was important to know if they were related in any way. The correlations between these measures are, therefore, presented in Table 1. The results show that test anxiety is unrelated to any of the other measures, indicating the statistical independence of the measures defining Anxiety and Motivation. The correlations between the ACE and GPR and anxiety are also not significant. Nonsignificant correlations between ACE scores and perceptual recognition tests also demonstrate that the variable of intelligence was successfully partialled out by the experimental procedures.

The only correlation reaching statistical significance is between the ACE and need Achievement. This reflects the fact that one's need for achievement influences performance on the ACE to a small but significant degree.

*Testing conditions.* By defining the criterion groups according to the selected measures, it was possible to obtain 72 subjects with 18 subjects in each of the four criterion groups (i.e., HA-HM, LA-LM, HA-LM, and LA-HM). Half of the subjects in each of these groups were randomly selected to take the tests under Stress conditions while the other subjects were given the tests under Nonstress conditions.

There were two basic variations in the testing situation under the Stress and Nonstress conditions. These variations were in the manner of conducting the testing, and the nature of the instructions given to the subjects.

The manner of conducting the testing session varied from one of formality under the Stress conditions to one of informality under the Nonstress conditions. In the Stress situation the experimenter was always dressed in a business suit, and he always introduced himself and addressed the subject by using the title "Mister," followed by the ap-

TABLE 1

CORRELATIONS BETWEEN MEASURES OBTAINED IN  
DEFINING CRITERION VARIABLES

	ACE	GPR <sup>a</sup>	Need Achievement	Anxiety
ACE				
GPR <sup>a</sup>	.03			
Need Achievement	.23*	.07		
Anxiety	.00	.04	-.04	

\* Correlations with GPR are biserial corrected for widespread groups, others are product-moment. All correlations are from 292 sophomore, male students at the University of Texas with ACE scores and GPRs as defined in text.

\* Significant beyond .001 level of confidence.

propriate surname. In general, it could be assumed that this situation was most characteristic of a student-instructor relationship.

Throughout the Nonstress situation the atmosphere fostered by the experimenter was more characteristic of a student-student relationship. Consequently, in this situation the experimenter always dressed in sportswear, always introduced himself by using only his first and last names, and always used the subject's first name when giving instructions and answering questions.

In addition to varying the general atmosphere of the testing situation, the instructions given the subjects in the two situations varied in a number of ways so as to act consistently with the changes in the atmosphere. For instance, subjects taking the tests under the Stress conditions were told that they were about to take a "new type of intelligence test" that was known to be highly related to grades and academic standing; they were kept in the dark as to the nature or composition of the test; and they were told that this test has so far proven to be quite easy for college students. In the Nonstress situation the subjects were told that the test they were about to take was relatively new and, therefore, "nobody knew what it measured, to what it might be related, or the meaning of the results." Furthermore, they were given a general description of the different parts of the test in addition to being told that some parts of the test were quite difficult and that no one was expected to be able to give the correct answers to all of the cards.

In setting up these variations in the testing situation, it was realized that it would be impossible to know which variables in the situations brought about a change in performance on the tests, if any changes were found at all. Such an analysis was not, however, the object of study. Instead, the primary purpose was to try to study the effect, if any, of changing the degree of threat or ego involvement for the subject as it may be related to the variables of Anxiety and Motivation.

*Perceptual recognition measures.* All subjects meeting the criteria defining the HA or LA and the HM or LM groups were individually given five perceptual recognition tests in the following order: Holtzman Figure-Recognition Test, achromatic form; Holtzman Figure-Recognition Test, chromatic form;



TABLE 2  
ANALYSIS OF VARIANCE FOR THE HOLTZMAN ACHROMATIC FIGURES

Source of variation	SS	df	MS	F
Anxiety (A)	748	1	748	
Motivation (M)	1,240	1	1,240	
Testing conditions (TC)	866	1	866	
A × M	6,613	1	6,613	12.36*
A × TC	122,584	1	122,584	9.38*
M × TC	94,033	1	94,033	
A × M × TC	589	1	589	
Between-subjects, within-groups	637,193	64	9,958	
Total between-subjects	863,867	71		
Between trial blocks (TB)	481,511	3	160,504	56.46*
TB × A	7,896	3	2,632	
TB × M	1,323	3	441	
TB × TC	5,712	3	1,904	
TB × Pooled conditions	30,045	12	2,504	
TB × Pooled subjects within groups	545,839	192	2,843	
Total	1,936,193	287		

\* Significant at .001 level of confidence.

Witkin's Embedded Figures Test; Hidden Digits Test; and Mooney Closure Test. All of these tests are similar in structure and in design. Essentially they are tests measuring skill in recognition of a definite animate or inanimate object "hidden" among an ambiguous configuration of perceptual cues that make up the background stimuli. The Holtzman tests are made up of animate figures hidden among a configuration of either chromatic or achromatic background stimulus cues. The other tests are described in detail by Thurstone (1944).

Reaction time scores were taken on all items of

each test. An inverse square root transformation of these scores was used in the statistical analysis on only four of the tests, however. On the Mooney Closure Test, where a number of the items were incorrectly identified within a 2-minute time limit, it was necessary to use a pass-fail criterion for scoring and a square root transformation in the statistical analysis.

From empirical studies, especially those done by Thurstone (1944), it is possible to gain some understanding of what is being measured by these tests. In his factorial study of perceptual tests, Thurstone found the original embedded figures, the hidden digits, and the original street-gestalt version of the Mooney Figures to have a common factor which he thought was measuring the ability to form a perceptual closure against distraction. In addition to having loadings on this factor, the Gottschaldt figures and the hidden digits were shown to have high loadings on a second factor which Thurstone thought represented flexibility in dealing with conflicting gestalts. These factors are commonly acknowledged as part of the spatial relations factors found on most intelligence tests.

Although the Holtzman Figures have not previously been correlated with other perceptual tests, they were constructed to measure perceptual closure against distracting and conflicting stimuli. Therefore, it would be expected that perceptual processes similar to those identified in the other tests in this battery are operating in this test. The correlations obtained in the process of carrying out this study support this conclusion. The intercorrelations of all these tests range from .41 to .70, all significant at the .001 level of confidence or better.

## RESULTS

The results of an analysis of variance for the Holtzman recognition test, achromatic

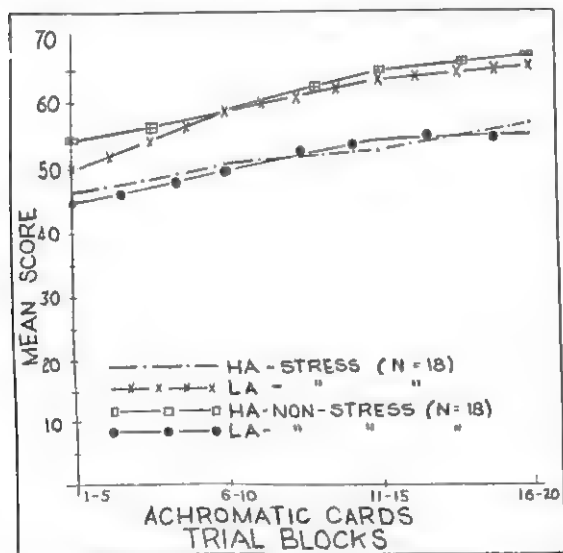


FIG. 1. Performance curves on the Holtzman Achromatic Cards for HA and LA groups under Stress and Nonstress conditions. (Curves are based on five-trial averages for each group. Additional markings are to aid in identifying each specific curve.)

version, are presented in Table 2. The findings for this test are exactly the same as the other tests with the one exception of the Hidden Digits Test where the  $F$  ratios for  $A \times TC$  and  $M \times TC$  only approach statistical significance at the .05 level of confidence. A low odd-even reliability coefficient of .58 for this test would seem to account for these lowered  $F$  ratios.

The performance curves presented in Figure 1 illustrate the specific interpretations that can be derived from the analysis of variance for the anxiety groups. Curves for the HM and LM groups are essentially replications of those shown in Figure 1. Thus we find that both Anxiety and Motivation, as defined by the criterion measures, function in the same way in each of the testing situations.

In Figure 1 the curves for the HA Stress group and the LA Nonstress group are not significantly different from each other. Both curves reflect a gradual but significant improvement in performance from trial to trial. Likewise, the curves for the LA Stress group and the HA Nonstress group are not significantly different from each other. Again both curves reflect a steady improvement in performance. Thus all groups show an improved performance across trials that is significant. However, there is no significant interaction between this practice effect and the variables of Anxiety, Motivation, or the testing conditions.

There are, however, very significant interactions between Anxiety and the testing conditions and between Motivation and the testing conditions. Furthermore, these interactions are in the same direction. Thus the LA Stress and the HA Nonstress groups have shorter reaction times than do the LA Nonstress and the HA Stress groups.<sup>4</sup> Likewise, the LM Stress and the HM Nonstress groups have shorter reaction times than do the LM Nonstress and the HM Stress groups. Both the LA and the LM groups have shorter reaction times than their respective HA and HM groups under the Stress conditions, while both the HA and the HM groups have shorter reaction times than their respective LA and LM groups under the Nonstress conditions.

<sup>4</sup> The use of inverse square root transformations accounts for the reversed appearance of the curves in Figure 1.

## DISCUSSION

### Anxiety

The significantly longer reaction times of the HA group over the LA group in the Stress situation are consistent with the belief that anxiety functions as a negative drive stimulus which impedes performance on a problem solving task. However, the fact that this difference remains from trial to trial is inconsistent with our predictions and with the results of Mandler and Sarason (1952) where they found that early differences between the HA and the LA groups disappeared gradually as reinforcement was obtained from trial to trial. These results appear to be contradictory to their original ideas. However, closer examination of the basic conditions underlying the interaction of Trial  $\times$  Anxiety found by Mandler and Sarason reveals that what these authors consider a Trial  $\times$  Anxiety interaction is probably a Situation  $\times$  Anxiety interaction. Thus during the early trials of their task, the situation is more threatening to the HA group than it is to the LA group; however, during the later trials the stressful aspects of the situation disappear with the subsequent disappearance of the difference in the levels of performance of the two anxiety groups.

In the present study, it seems that the stress or nonstress aspects of the testing conditions remained relatively constant throughout each test as a result of the deliberate attempt to create and maintain either a competitive and threatening atmosphere or a noncompetitive and relaxed atmosphere. However, in the Mandler and Sarason (1952) study, the testing conditions were designed to be more neutral; and once the threat presented by the novelty of their situation subsided, the improved performance was manifested by the HA group.

Although the inferior performance of the HA over the LA subjects in the Stress situation is as predicted, the fact that the HA group has significantly faster reaction times than the LA groups when the tests are given in the Nonstress situation represents a complete reversal of our prediction for the Nonstress conditions. This is in sharp contrast to stress conditions. This is in sharp contrast to the results in the Stress situation where some differential, negative drive force is operating.

These positive and negative drive aspects of anxiety are not basically unlike those originally postulated by Mandler and Sarason (1952). Instead of finding anxiety to change functional characteristics across trials, it is necessary to consider anxiety as having different functional characteristics according to the Stress, Nonstress features of the testing situation.

Two studies by Sarason (1956, 1957a) have demonstrated similar functional characteristics of anxiety. Using college students he found that "motivating instructions" detrimentally affected the performance of HA subjects while LA and MA subjects were facilitated on a serial learning task. In another very similar experiment testing the retention of nonsense syllables, it was found that HA subjects were significantly superior to LA subjects given neutral instructions while the opposite results were obtained under so-called subject oriented instructions emphasizing "the need for S to perform well in order to maintain his self-esteem" (Sarason, 1957a, p. 171).

In another study by Ruebush (1960) on sixth grade boys with Otis beta IQ scores between 111 and 122, it was found that the HA subjects were inferior to LA subjects on a short form of the Witkin's Embedded Figures Test (Jackson, 1956). In analyzing each subject's approach to the testing situation by means of a cautiousness index, the author concludes that the differential "effect of anxiety on performance . . . is mediated primarily by defensive reactions to the anxiety"—and that these effects "may vary systematically depending upon both intelligence level and type of task and instruction" (Ruebush, 1960, p. 211).

These studies, along with the present results, indicate that anxiety can function *both* as a positive or as a negative drive on the same task depending on the stress, nonstress characteristics of the situation in which the tasks are completed. This being the case, it would seem that anxiety must be considered not only in terms of the type of response involved, but also in terms of the *task situation*.

Some explanation must be found for the "reversal effect," namely, that HA subjects do more poorly under stress than under lack of stress, whereas the LA subjects do better un-

der stress than under lack of stress. One interpretation is that, up to a moderate level of intensity, anxiety is directly related to the level of performance; however, once past this threshold, anxiety is inversely related to level of performance. Under such conditions, it would be expected that when anxiety is at a medium level subjects would perform at their highest level while it would be expected that under conditions where anxiety is at a very low intensity or is at a very high intensity, subjects would perform at their lowest level. In that the LA group under Stress and the HA group under Nonstress performed significantly better than did the LA group under Nonstress and the HA group under Stress, these expectations are consistent with the results of this study.

The results from several studies reported by Matarazzo and his colleagues are also consistent with this interpretation. In two studies using college students (Matarazzo & Phillips, 1955; Matarazzo, Ulett, & Saslow, 1955), one involving maze learning, the other using a digit symbol test, the MA groups performed on a consistently higher level than the HA and LA groups. In a third study using Veterans Administration inpatients on a pursuit-motor task (Matarazzo & Matarazzo, 1956), similar trends were observed.

The concept of "activation" or "energy mobilization" as developed by Duffy (1957) seems to be applicable to this interpretation. This concept "is designed to describe the intensity aspect of *all* behavior" and is defined as "the extent to which the organism as a whole is activated" (p. 266). Although this concept refers to the degree of internal arousal, Duffy acknowledges the possibility of a discrepancy between the degree of arousal and the overt response, a discrepancy which she believes to be due to the intervention of inhibitory processes. These inhibitory processes could be conceived of in terms of the competing or irrelevant responses of Spence and Sarason. Furthermore, Duffy also believes that the degree of optimal activation appears to be a moderate one, the curve expressing the relationship between activation and quality of performance taking the form of the inverted U. Thus the disorganization of responses during overstimulation (e.g., our



HA subjects under stress) would represent one end of this U shaped curve.

### Motivation

Since Mandler and Sarason (1952) tried to integrate their ideas about the functioning of anxiety with those of a task oriented drive, an attempt was made to define this variable and predict its functional characteristics. Thus it was predicted that HM subjects would have significantly faster reaction times than LM subjects. This prediction is supported only under Nonstress conditions. Under Stress conditions, the same "reversal effect" appears, since the LM group is found to have significantly faster reaction times than the HM group. Thus the results testing the effects of stress on this drive variable are identical to those of anxiety.

In the light of the independence of the operational definitions of motivation and anxiety, the question arises: why the functional equivalence of these two drives? The conceptions concerning the functional properties of motivation as conceived in this study must be considered inadequate. The evidence suggests that a drive variable that is defined independently of anxiety can also have both positive and negative functional characteristics on the same task. This would be expected, if Duffy's (1957) concept of activation were applicable to different drive variables.

### SUMMARY

Seventy-two male, sophomore, college students with middle range scores on the ACE were selected on the basis of operationally defined criteria as being either high or low in Anxiety and either high or low on an independent Motivation variable. A series of five perceptual recognition tests were administered individually to all subjects. Half of the subjects in each group were given the tests under conditions designed to be stressful, while half the subjects in each group were given the tests under Nonstress conditions.

In the Stress situation, the HA and the HM groups performed significantly less well than the LA and the LM groups, respectively. In the Nonstress situation, however, the LA and the LM groups performed significantly less well than the HA and the HM groups, re-

spectively. The weight of evidence from this and many other studies bearing on this problem points to the need to include the task and situational variables as a crucial factor in formulating a comprehensive theory of motivation.

### REFERENCES

- DUFFY, ELIZABETH. The psychological significance of the concept of "arousal" or "activation." *Psychol. Rev.*, 1957, **64**, 265-275.
- GORDON, E. M., & SARASON, S. B. The relationship between test anxiety and other anxieties. *J. Pers.*, 1955, **23**, 317-323.
- JACKSON, D. N. A short form of Witkin's Embedded Figures Test. *J. abnorm. soc. Psychol.*, 1956, **53**, 254-255.
- MANDLER, G., & COWEN, JUDITH E. Test Anxiety Questionnaires. *J. consult. Psychol.*, 1958, **22**, 228-229.
- MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166-173.
- MATARAZZO, J. D., & PHILLIPS, J. S. Digit symbol performance as a function of increasing levels of anxiety. *J. consult. Psychol.*, 1955, **19**, 131-134.
- MATARAZZO, J. D., ULETT, G. A., & SASLOW, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, 1955, **53**, 79-95.
- MATARAZZO, R. G., & MATARAZZO, J. D. Anxiety level and pursuit motor performance. *J. consult. Psychol.*, 1956, **20**, 70.
- ROCKETT, F. C. Form recognition as a function of test-anxiety, color figure, and color-pattern. Unpublished doctoral dissertation, University of Texas, 1955.
- RUEBUSH, B. K. Interfering and facilitating effects of test anxiety. *J. abnorm. soc. Psychol.*, 1960, **60**, 205-212.
- SARASON, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253-260.
- SARASON, I. G. Effect of anxiety and two kinds of motivating instructions on verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 166-171. (a)
- SARASON, I. G. Test anxiety, general anxiety, and intellectual performance. *J. consult. Psychol.*, 1957, **21**, 485-490. (b)
- SARASON, I. G. Intellectual and personality correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1959, **59**, 272-275.
- SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810-817.
- THURSTONE, L. L. *A factorial study of perception*. Chicago: Univer. Chicago Press, 1944.
- WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, PEARL B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

(Received February 14, 1961)

## SIGNIFICANT FACTORS IN HYPNOTIC BEHAVIOR<sup>1</sup>

THEODORE XENOPHON BARBER AND LOUIS BENJAMIN GLASS<sup>2</sup>

*State Hospital, Medfield, Massachusetts*

Wells' (1924) informal studies on "waking hypnosis" suggest the following hypotheses:

1. Direct suggestions, such as "Your body is immovable," or "You cannot say your name," administered *without* a preliminary trance induction procedure, are sufficient to produce body immobility, verbal inhibition, and other behaviors historically associated with the word "hypnosis" in a "considerable proportion" of subjects.

2. Such "hypnotic behaviors" can be as effectively elicited by direct suggestions given alone as by suggestions given after a formal trance induction.

The results of the study of Weitzenhoffer and Sjöberg (1961) tend to support the first hypothesis but not the second. These investigators found that suggestions given after a formal induction procedure were more effective than the same suggestions given alone in producing arm rigidity, eye catalepsy, visual-auditory hallucination, negative hallucination, and other behaviors termed "hypnotic." There is additional confirmation of the first hypothesis in the recent studies of Barber (1959), Barber and Deeley (1961), and Barber (1961).

The purpose of the present investigation was twofold: to provide a further test of Wells' (1924) hypotheses (Studies I and II) and to explore the possible biographical and attitudinal factors which may be associated with responsiveness to direct suggestions (Study III).

### STUDY I. PARAMETERS OF "SUGGESTIBILITY" IN A STUDENT POPULATION

#### *Method*

*Subjects.* All subjects resided in college dormitories and were obtained by a room-to-room ap-

<sup>1</sup> This investigation was supported by a research grant (MY3253) from the National Institute of Mental Health, of the National Institutes of Health, United States Public Health Service.

The authors wish to thank Joan Burke for administering the tests in Study I and Milton S. Palanker for obtaining the male subjects.

proach in which they were asked to participate in an experiment on imagination and concentration (hypnosis was not mentioned). Of 497 students approached, 462 (93%) agreed to participate. The 462 participants consisted of 236 women at Regis College, and 123 women and 103 men at Boston University. The age range was 17-28 with a median of 19.

*Procedure.* All subjects were tested by the same experimenter.<sup>3</sup> Preliminary instructions were:

These are all tests of imagination. The better you can imagine and the harder you try, the more you'll respond. Try as hard as you can to concentrate, and to imagine the things I tell you.

As an introduction to the tasks, the Chevreul Pendulum Test, as described by Weitzenhoffer (1953, p. 12), was administered. Each subject was then seated in a comfortable chair and given the following eight tests (direct suggestions) in the same way, as memorized by the experimenter; the first three tests were given with the subject's eyes open, the remaining five with eyes closed.

1. Arm Lowering. Starting with the subject's right arm extended and horizontal, suggestions are given for 30 seconds that the arm is becoming heavy and is moving down. Scoring criterion: 1 point for response of 4 inches or more.

2. Arm Levitation. Starting with the subject's left arm extended and horizontal, suggestions are given for 30 seconds that the arm is weightless and is moving up. Scoring criterion: 1 point for response of 4 inches or more.

3. Hand Lock. Subject is instructed to clasp his hands together tightly with fingers intertwined and place them in his lap. Suggestions are given for 45 seconds that the hands are like steel, they are welded together, they cannot be taken apart. Scoring criteria:  $\frac{1}{2}$  point for incomplete separation of hands after 5-seconds effort; 1 point for incomplete separation after 15-seconds effort.

4. Thirst "Hallucination." Suggestions of extreme thirst are given for 45 seconds. Scoring criteria:  $\frac{1}{2}$  point if the subject shows noticeable swallowing; moistening of lips, or marked mouth movements; additional  $\frac{1}{2}$  point if the subject indicates (during the post-experimental interview) that he became thirsty during this test.

5. Verbal Inhibition. Suggestions are given for 45 seconds that the subject's throat and jaw muscles are rigid and he cannot say his name. Scoring criteria:  $\frac{1}{2}$  point if the subject does not say his name after 5-seconds effort; 1 point if he does not say his name after 15-seconds effort.

<sup>2</sup> Now at the University of Michigan.

<sup>3</sup> Joan Burke.





(50%) passed from two to five tests, and the remaining 129 subjects (28%) passed less than two tests.

The number of subjects passing (scoring 1 point) on each test was: Arm Lowering—270 (58%), Arm Levitation—226 (49%), Hand Lock—96 (21%), Thirst "Hallucination"—277 (60%), Verbal Inhibition—69 (15%), Body Immobility—45 (10%), "Posthypnotic" Response—208 (45%), Selective Amnesia—63 (14%). If 45% represents a "considerable proportion" of subjects, the first hypothesis—that behaviors associated with the word "hypnosis" can be induced in a "considerable proportion" of subjects by direct suggestions without a preliminary trance induction procedure—holds, in the present sample, for arm lowering, arm levitation, thirst "hallucination," and "posthypnotic" response. Although suggestions of verbal inhibition, body immobility, and selective amnesia are rated on some scales of "hypnotic depth" (cf. LeCron & Bordeaux, 1949, pp. 65-66) as "difficult" suggestions which indicate a "medium" or "deep trance," a small minority of subjects (10-15%) in the present sample passed these tests without a formal trance induction.

The 60 subjects given a retest showed lower scores in the second session: the mean score on the first testing (3.13) is significantly higher than the retest score (2.58) ( $t = 3.35$ ,  $p < .01$ ). A product-moment correlation coefficient of 0.88 ( $p < .001$ ) was obtained between scores in the two sessions.

The results of this investigation are in line with Weitzenhoffer and Sjöberg's (1961) study which found that "some individuals can produce hypnotic-like behavior in the absence of any formal induction of hypnosis." Previous studies from our laboratory suggest a similar conclusion. Barber and Deeley (1961) found that some un hypnotized persons who are instructed to "try not to see the color red and the color green" give responses to the Ishihara Plates similar to those given by "hypnotic color-blind" subjects. Oswald (1957) and Barber (1959) have presented data indicating that a performance considered characteristic of the "deeply hypnotized" person—"hallucinating" colors on a gray card, and subsequently reporting ap-

propriate complementary colored afterimages (Erickson & Erickson, 1938; Rosenthal & Mele, 1952)—can be elicited from some persons (without a preliminary trance induction) by instructions to imagine the specified colors. Additional investigations, recently reviewed elsewhere (Barber, 1961) suggest that a trance induction procedure is not necessary to elicit other behaviors associated with the term "deep hypnosis" such as hypnotic deafness, hypnotic blindness, and negative hallucinations; it appears likely that similar performances can be induced in some normal persons by the following brief instructions:

Try to remain inattentive to the auditory (or visual) stimuli.

## STUDY II. THE RELATIVE EFFECTIVENESS OF DIRECT SUGGESTIONS AND A CONVENTIONAL TRANCE INDUCTION IN ELICITING "HYPNOTIC BEHAVIORS"

The present study was designed to test the second hypothesis: Direct suggestions given alone are as effective as the same suggestions given after a formal trance induction in eliciting behaviors associated with the word "hypnosis."

### Method

*Subjects.* Thirty subjects (20 men and 10 women), who had not been previously hypnotized, volunteered to participate. Fourteen were male college students and 16 were adults residing at a summer camp. The age range was 18-41 with a median of 24.

*Procedure.* Subjects were informed in advance that two experimental sessions were planned, the first session involving only tests of imagination given under normal conditions, and the second session involving hypnosis.

Instructions for the first session (control) were:

Today you will be given tests of imagination under normal conditions. You will not be hypnotized. The better you can imagine and the harder you try, the more you'll respond; try as hard as you can to concentrate, and to imagine the things I tell you.

The seated subject was instructed to close his eyes and the eight tests (direct suggestions) were administered as described in Study I above.<sup>5</sup>

<sup>5</sup> In this experiment, Form A of Test 7 and 8, as described in Footnote 4, was used, and the Chevreul Pendulum was not employed as a preliminary test.

The second session (formal trance induction) was held after a lapse of not less than two days or more than one week. Preliminary instructions were:

In the previous session, I tested your ability to imagine under normal conditions; today you're going to be deeply hypnotized and I think you'll find it an interesting and enjoyable experience.

The seated subject was instructed to fixate on an electric metronome which ticks in synchrony with a blinking light, and a 20-minute trance induction, patterned after the procedures outlined by Friedlander and Sarbin (1938), Weitzenhoffer and Hilgard (1959), and Marcuse (1959) was administered by an experimenter (LBG) with prior training as a hypnotist. Immediately after the trance induction procedure, the eight tests were administered (as if they were part of the induction procedure) and scored in the same way as in the preceding (control) session.

### *Results and Discussion*

The test-retest results from Study I indicate that if the eight tests are administered twice (without a preliminary trance induction) the mean score tends to drop (by about 0.5 points) in the second session; in the present study, with a formal trance induction procedure preceding the tests in the second session, the average score not only did not drop the second time but was higher (by 0.5 points) than the average score in the first session. A comparison of the scores under the control and experimental conditions using Wilcoxon's matched-pairs signed-ranks test showed that the latter were significantly higher than the former ( $T = 33.5$ ,  $N = 19$ ,  $p < .02$ ).<sup>6</sup> The second hypothesis thus appears untenable: suggestions given after a conventional trance induction procedure are more effective than the same suggestions given alone in eliciting behaviors associated with the word "hypnosis."

Although the trance induction procedure was associated with a higher mean score, 4 subjects (13%) attained higher scores under the control condition (first session), 11 (37%) attained the same score under the two conditions, and only 15 subjects (50%) showed higher scores after the trance induction (second session). Of the 15 subjects at-

taining higher scores under the experimental condition, 7 gained no more than 1 point and 8 (or 27%) showed larger gains (1.5-3 points). A Spearman rank-correlation coefficient of 0.85 ( $p < .001$ ) was obtained between scores in the two sessions.

Weitzenhoffer and Sjöberg (1961) previously reported similar findings. In their experiment, 17 suggestions of graded difficulty were administered to 30 college students in the "waking" condition; when a subject had failed three successive suggestions, a 15-minute trance induction procedure (consisting of eye-fixation and suggestions of relaxation, drowsiness, and sleep) was administered. After the trance induction, the tests were resumed, beginning two items below the first item failed and continuing until the scale was completed or the subject had again failed three successive items. The scores on the 17-point "suggestibility" scale were significantly higher following the trance induction. However, 17 subjects (57%) did not gain in score after the induction and, of the 13 subjects (43%) showing higher scores, only 5 (17%) showed marked gains (5 or more points on the 17-point scale). In brief, the results of the present experiment (Study II) and the results of the experiment previously reported by Weitzenhoffer and Sjöberg are in agreement as follows: (a) a conventional trance induction procedure is associated with a statistically significant overall enhancement of "suggestibility" in a group of subjects, (b) this overall gain is due to higher scores attained by approximately half of the subjects, and (c) only a small proportion of subjects (about one out of five) show relatively large gain.

The trance induction procedure employed in the present experiment (Study II) included the following components: (a) The situation was explicitly defined as hypnosis and implicitly defined as different, as important, as a situation in which maximal responsiveness and unusual behavior was expected. The subject (b) fixated on a blinking light while listening to the repetitive tick of the metronome. (c) remained relatively immobile with eyes closed for 20 minutes, and (d) listened to the hypnotist's statements to the effect that he was becoming relaxed, drowsy, and

<sup>6</sup> Application of the Wilcoxon signed-ranks test to the eight tests individually showed no significant difference between scores under experimental and control conditions on any one test.



sleepy. Could heightened response to suggestions have been obtained if the preliminary procedure was limited only to the first of these components—i.e., if the situation was defined either as hypnosis or as a very important test situation in which behavior normally considered odd or inappropriate was desired and expected—without employing Components *b*, *c*, and *d*? An affirmative answer to this question is suggested by a follow-up experiment (Glass & Barber, 1961) which appears to indicate that, with at least some subjects, a placebo administered by a physician as a "hypnosis producing" drug is as effective as a formal 20-minute trance induction procedure in enhancing response to the eight tests (direct suggestions) described in Study I above. Further experiments are in order to delineate which of the factors subsumed under the term "trance induction procedure" are necessary, which helpful but unnecessary, and which superfluous to producing enhanced response to suggestions.

### STUDY III. AN EXPLORATION OF PERSONALITY CHARACTERISTICS RELATED TO "SUGGESTIBILITY"

Our recent experimental studies (Barber, 1956, 1957a, 1957b, 1958, 1960) and a series of earlier investigations reviewed by Weitzenhoffer (1953) suggest a possible relationship between the characteristics of the subject, as a person, and responsiveness to suggestions of the type commonly employed in hypnosis experiments. This indicated relationship was explored in the present study.

In a prior investigation with 70 subjects (Barber, 1960), 61 items from three questionnaires (the Guilford-Zimmerman Temperament Survey, the Webster-Sanford-Freedman (1955) version of the F Scale, and a 14-item questionnaire especially constructed for the study) were found to differentiate "highly suggestible" from "unsuggestible" subjects. A questionnaire consisting of these 61 items was administered to the three groups participating in Study I (women at Regis College, women at Boston University, and men at Boston University), and to the single group participating in Study II. Analysis of the results for the 106 "highly suggestible" and the 140 "unsug-

gestible" subjects<sup>7</sup> revealed four items which differentiated the latter from the former in all four groups and which were significant at the .02 level by the chi square test. To determine if these four items would "hold-up" in a previously untested group, an additional group (20 women college students) was given the questionnaire and rated on the eight tests (direct suggestions) as described in Study I (by the same experimenter who had carried out Study I). The three "highly suggestible" and the six "unsuggestible" subjects in the new group answered each of the four items in the predicted direction. The number of positive items, the content of the items, and the fact that each item held up over five groups tested by two experimenters is sufficient to exclude chance findings (Brozek & Tiede, 1952). The differentiating items are as follows ("highly suggestible" subjects tend to answer Yes; "unsuggestible" subjects tend to answer No):

1. You like to sell things (that is, act as a salesman) ( $\chi^2 = 7.08$ ,  $p < .01$ ).
2. You like to read true stories about love and romance ( $\chi^2 = 11.07$ ,  $p < .001$ ).
3. You find daydreaming very enjoyable ( $\chi^2 = 6.05$ ,  $p < .02$ ).
4. When you were a child of about five or six, did you have imaginary playmates who were rather vivid and almost real? ( $\chi^2 = 20.86$ ,  $p < .001$ ).

In a follow-up interview, the "highly suggestible" subjects (in Study I) who had given a Yes answer to Item 1 were asked, "Why do you like to sell things (that is, act as a salesman)?" the answers obtained (e.g., "Because I like people," "I like to meet new people") suggest that Item 1 refers to readiness to form new interpersonal relationships. Items 2-4 appear to refer to a pattern of behavior which can be provisionally categorized as proneness to imaginative activities, daydreaming, and fantasy.

Although the four items "held-up" over the five groups tested in these experiments, the following should be noted:

<sup>7</sup> Subjects attaining scores of 5.5 or above on the eight tests (direct suggestions) are categorized as "highly suggestible"; subjects scoring below 2 are categorized as "unsuggestible."

One of the "highly suggestible" subjects ( $N = 107$ ) and one of the "unsuggestible" subjects ( $N = 141$ ) did not complete the questionnaire.



1. Only two experimenters participated and, with the exception of 16 subjects tested in Study II, all subjects were college students; it remains to be determined if the items apply to nonstudent subjects and to subjects tested by different experimenters.

2. The items differentiate only the "highly suggestible" from the "unsuggestible" subjects and do not apply to the large group rated medium on "suggestibility."

3. Since only 4 of the original 61 items "held" over the five groups tested by two experimenters, it appears possible that "suggestibility" is not in one-to-one correspondence with fixed "personality characteristics" but is multidetermined—by situational factors, by the characteristics of the experimenter, by an interaction between the subject's characteristics and situational factors, and by an interaction between the subject's characteristics and the experimenter's characteristics.

#### SUMMARY

The primary purpose of this investigation was to test two hypotheses suggested by hypnosis": (a) Direct suggestions (administered without a preliminary trance induction procedure) are sufficient to elicit behaviors associated with the word "hypnosis" from a "considerable proportion" of subjects, and (b) behaviors associated with the word "hypnosis" can be as effectively elicited by direct suggestions given alone as by suggestions given after a formal trance induction procedure.

In an investigation (Study I) concerned with the first hypothesis, three groups of male and female college students ( $N = 462$ ) were given eight direct suggestions designed to elicit the following items of behavior: arm lowering, arm levitation, inability to unclasp hands, thirst "hallucination," inability to say name, body immobility, selective amnesia, and "posthypnotic" response. Although a formal trance induction procedure was not employed, 102 subjects (22%) carried out more than five of the eight suggested behaviors. The experimental behaviors and post-experimental reports of these subjects were comparable to what one might expect had they been formally "hypnotized." An addi-

tional 231 subjects (50%) responded to between two to five suggestions, and the remaining 129 subjects (28%) responded to less than two suggestions. If 45% represents a "considerable proportion" of subjects, the first hypothesis was confirmed, in this sample, for the following behaviors: arm lowering, arm levitation, thirst "hallucination," and "posthypnotic" response.

In an experiment (Study II) designed to test the second hypothesis, 30 subjects were given the eight direct suggestions twice: first, without a preliminary trance induction procedure (control condition) and, second, following a formal trance induction (experimental condition). The second hypothesis was not confirmed: in harmony with a previous study reported by Weitzenhoffer and Sjöberg (1961), it was found that a trance induction procedure had a significant overall effect in enhancing response to suggestions. Fifteen subjects (50%) attained higher scores under the experimental condition, 11 (37%) attained the same score under the two conditions, and 4 (13%) attained higher scores under the control condition.

A concomitant study (Study III) was carried out to explore an indicated relationship between personality characteristics and responsiveness to suggestions of the type commonly employed in hypnotic experiments. A questionnaire constructed in an earlier investigation (Barber, 1960) was administered to the three groups rated on "suggestibility" in Study I, to the group rated in Study II, and to an additional group of 20 subjects rated in a follow-up study. The results suggest that in this sample of American college students the "highly suggestible" subjects, as compared with the "unsuggestible," are more willing to form new interpersonal relationships, and are more prone to imaginative activities, daydreaming, and fantasy.

#### REFERENCES

- BARBER, T. X. A note on "hypnotizability" and personality traits. *J. clin. exp. Hypnosis*, 1956, 4, 109-114.  
 BARBER, T. X. Experiments in hypnosis. *Scient. American*, 1957, 196(4), 54-61. (a)  
 BARBER, T. X. Hypnosis as perceptual-cognitive restructuring: I. Analysis of concepts. *J. clin. exp. Hypnosis*, 1957, 5, 147-166. (b)

- BARBER, T. X. The concept of "hypnosis." *J. Psychol.*, 1958, **45**, 115-131.
- BARBER, T. X. The after-images of "imagined" and "hallucinated" colors. *J. abnorm. soc. Psychol.*, 1959, **59**, 136-139.
- BARBER, T. X. The necessary and sufficient conditions for hypnotic behavior. *Amer. J. clin. Hypnosis*, 1960, **3**, 31-42.
- BARBER, T. X. Physiological effects of "hypnosis." *Psychol. Bull.*, 1961, **58**, 390-419.
- BARBER, T. X., & DEELEY, D. C. Experimental evidence for a theory of hypnotic behavior: I. "Hypnotic color-blindness" without "hypnosis." *Int. J. clin. exp. Hypnosis*, 1961, **9**, 79-86.
- BROZEK, J., & TIEDE, K. Reliable and questionable significance in a series of statistical tests. *Psychol. Bull.*, 1952, **49**, 339-341.
- ERICKSON, M. H., & ERICKSON, ELIZABETH, M. The hypnotic induction of hallucinatory color vision followed by pseudo-images. *J. exp. Psychol.*, 1938, **22**, 581-588.
- FRIEDLANDER, J. W., & SARBIN, T. R. The depth of hypnosis. *J. abnorm. soc. Psychol.*, 1938, **33**, 281-294.
- GLASS, L. B., & BARBER, T. X. Hypnotic behavior, the definition of the situation, and the placebo-effect. *J. nerv. ment. Dis.*, 1961, **132**, 539-541.
- LECROON, L. M., & BORDEAUX, J. *Hypnotism today*. New York: Grune & Stratton, 1949.
- MARCUSE, F. L. *Hypnosis—fact and fiction*. Baltimore: Penguin, 1959.
- OSWALD, I. After-images from retina and brain. *Quart. J. exp. Psychol.*, 1957, **9**, 88-100.
- ROSENTHAL, B. G., & MELE, H. The validity of hypnotically induced color hallucinations. *J. abnorm. soc. Psychol.*, 1952, **47**, 700-704.
- WEBSTER, H., SANFORD, N., & FREEDMAN, M. A. new instrument for studying authoritarianism in personality. *J. Psychol.*, 1955, **40**, 73-84.
- WEITZENHOFFER, A. M. *Hypnotism—an objective study in suggestibility*. New York: Wiley, 1953.
- WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford hypnotic susceptibility scale*. Palo Alto, Calif.: Consulting Psychologists Press, 1959.
- WEITZENHOFFER, A. M., & SJOBERG, B. M., JR. Suggestibility with and without "induction of hypnosis." *J. nerv. ment. Dis.*, 1961, **132**, 204-220.
- WELLS, W. R. Experiments in waking hypnosis for instructional purposes. *J. abnorm. soc. Psychol.*, 1924, **18**, 389-404.

(Received February 23, 1961)

## CRITIQUE AND NOTES

### THE RELATIONSHIP BETWEEN RESPONSE STYLE AND PERSONALITY VARIABLES:

#### I. THE MEASUREMENT OF RESPONSE ACQUIESCENCE<sup>1</sup>

RICHARD K. MCGEE<sup>2</sup>

*Moccasin Bend Psychiatric Hospital, Chattanooga, Tennessee*

A recent approach to the response set component of test scores is to treat it not as error variance, but as a personal stylistic variable of interpretive significance in personality assessment (Jackson & Messick, 1958). Much attention has been focused on *response acquiescence*, the tendency to agree with test items, regardless of their content. Couch and Keniston (1960) presented the most recent data to confirm the general proposition that the "agreeing response tendency is based on a central personality syndrome" (p. 173). Typically, this conclusion is reached after the demonstration of a relationship between a single measure of acquiescence and some measure of conformity, authoritarianism, rigidity, dependency, anal resentment, or low ego strength, to name only a few. McGee (1962) has reviewed these studies elsewhere, and discussed some methodological revisions for the design of future investigations.

It may be observed that each of the major investigations of the personality correlates of response acquiescence has used a different instrument for measuring the response style. It would be valuable to know how much variance, if any, these measures have in common. To hypothesize that response style has its roots in the responder's personality implies an assumption that the response tendency is a generalized behavior which may be expected to appear in a variety of "agree-disagree" situations. Unless this assumption can be verified, it would be difficult to defend the position that a personal mode of responding springs from a central personality syndrome.

There is general agreement in the literature that the verbal content of items is relatively unimportant; that acquiescent individuals respond "Yes," or "Agree" irrespective of item content. Jackson and Messick (1958) reviewed the evidence, and concluded that "the major common

factors in personality inventories of the true-false or agree-disagree type . . . are interpretable primarily in terms of style rather than specific item content" (p. 247). What is important is the degree of structure, or ambiguity, of the stimuli. Cronbach (1946) stated that response sets are most likely to appear when the stimuli are ambiguous. Berg and Rapaport (1954) confirmed this prediction by showing that an "unstructured questionnaire," in which subjects responded to imaginary items, yielded a consistent tendency for acquiescent response options to be selected.

The foregoing considerations suggest the major hypothesis which it is the purpose of this study to test: Response acquiescence is a general test taking behavior which appears in any situation characterized by the forced choice of "agree-disagree" response options. This hypothesis may be examined by observing the appearance of response acquiescence when specific variations are made in the stimulus conditions, i.e., the item content. For example, if stylistic determinants, rather than content determinants account primarily for the variance of test scores, high correlations should be found between two measures of acquiescence, even though they differ markedly from one another in the meaningfulness of their verbal content. Similarly, if response styles appear most readily in ambiguous situations, high correlations should be found between measures of acquiescence which contain highly ambiguous stimuli.

#### METHOD

Six tests of acquiescence were administered to 218 undergraduate students in general psychology classes. The sample included 144 males and 74 females, ranging in age from 17 to 24 years. Five of the tests were taken in large group sessions approximating 100 subjects each, and lasting about 1 hour. The sixth test was presented to groups of about 40 subjects each in 20-minute sessions.

For the larger sessions, the five tests were mimeographed and assembled into booklets. Order of presentation was varied by using 10 random arrange-

<sup>1</sup> The author is indebted to Samuel S. Komorita and Jum C. Nunnally for their aid in planning this study and preparing this report of the results.

<sup>2</sup> This investigation was completed while the author was at Vanderbilt University.



ments, and subjects were told to complete the forms as they appeared in their booklet. The sixth test consisted of stimulus pictures which were projected by an opaque projector onto a screen at the front of the classroom. This form was taken by approximately half of the subjects prior to the group of five tests, and afterwards by the remaining half.

The six tests used in this study were selected because they represent the major techniques used thus far in the research on acquiescence. Three of them contain items loaded with meaningful verbal content, whereas the other three are essentially content free. The content tests used were the Social Acquiescence Scale (SAS), the California F Scale, and the Agreement Response Scale (ARS). Bass (1956) constructed the 56-item SAS out of proverbs, aphorisms, and other famous sayings. The F Scale (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950) is well known for its use as an attempt to measure both acquiescence and authoritarianism. The 30-item Form 40-45 was used here. The ARS was published by Couch and Keniston (1960) as the best short form of their 360-item test of the agreeing response tendency. Each of the 15 items correlates .40 or better with the general agreement factor in their data.

The three noncontent tests used have been devised for the purpose of measuring "pure" acquiescence by forcing either agreement or disagreement responses to unstructured, or meaningless stimuli.<sup>3</sup> The subjects were told they were participating in three different experiments: A study of the recognition value of German words, a study of the effectiveness of Extra-Sensory Perception, and a study of the amount of information communicated by abstract art. The German Language Recognition Form (Nunnally & Husek, 1958) contains 24 sentences consisting of two German nouns coupled by English verbs and modifiers. The ESP scale consisted of 25 partial sentences which included only two- or three-word stems, an indication of the length of the sentence, and the final word. It was presented as a model of the classic ESP studies in which the real, complete sentences were "locked up in a safe downstairs." The subjects were to concentrate on the (imaginary) list and attempt to read enough of the statement to indicate their agreement or disagreement with it. The Ambiguous Pictures scale consisted of 20 meaningless designs each drawn with black india ink on pieces of white paper 8 inches square. Beneath each

of these designs was printed the title of a popular mystery story. The subjects were told this was a test of the information transmitting properties of abstract art.

For all of the noncontent tests the instructions urged that even though the items appeared silly, or meaningless, the responses would be useful; that the desired experimental effects could be reliably measured if the subjects indicated whether they agreed or disagreed with the statements (German sentences and ESP sentences) or with the choice of title for the pictures; and it was suggested that subjects should obey their first impulse in responding.

Agreement responses were recorded in the same manner for all six of the tests. Mimeographed answer sheets were distributed which contained 60 seven-point rating scales. The subjects checked their degree of agreement with each item according to the following anchors: Completely Disagree, Mostly Disagree, Slightly Disagree, Neutral, Slightly Agree, Mostly Agree, and Completely Agree. Instructions explicitly stated that the Neutral category was to be used only rarely, if at all; that a decision should be forced in one direction or the other. Each of the tests was scored by summing the scale values of the ratings over all the items.

## RESULTS

Table 1 shows the matrix of coefficients obtained for the six acquiescence scores and sex. Scatter diagrams indicated that the assumption of linear regression was not violated in the data, thus the product-moment formula was applied for the interscale correlations. The degree of relationship between the tests and sex was computed by the point biserial formula such that a positive correlation indicates that the mean for males is greater than the mean for females.

Corrected split-half (odd-even) reliability coefficients are shown in parentheses along the diagonal. It should be noted that the reliability of the SAS is within the range of values (.81-.92) which Bass (1956) reported for the instrument. The reliability estimate for the F Scale falls slightly below the .81-.89 range reported by Adorno et al. (1950) for the combined Form 40-45. However, variations in the size and other characteristics of the sample must be considered in making such a comparison. The important point is that only two of the six acquiescence scores appear to possess adequate reliability even for research instruments.

An important finding regarding the reliability of noncontent scales appears in connection with the German Language Form. Nunnally and Husek (1958) found substantial correlations (—.69 for example) between these same items and variables such as age and education. Therefore, the scale must have considerably greater reliability when

<sup>3</sup> Copies of the three noncontent test stimuli may be obtained without cost from the author. The ESP scale is similar in format to the "unstructured questionnaire" used by Berg and Rapaport (1954), however the present form of this technique was developed independently by Nunnally. The Ambiguous Pictures were constructed from the designs which form one subtest of Yacorzynski's battery for detecting brain damage, and from random combinations of items in the PRT (Berg, Hunt, & Barnes, 1949). These materials were available through the assistance of K. Edward Renner of Northwestern University, for which the author is grateful.

TABLE 1  
INTERCORRELATIONS BETWEEN SIX MEASURES OF ACQUIESCENCE AND SEX  
(*N* for all variables = 218)

Variable	SAS	F Scale	ARS	German words	ESP	Ambiguous pictures	Sex
SAS	(.85)	.56	.51	.27	.27	.01	-.06
F Scale		(.77)	.39	.06	.19	.01	.10
ARS			(.54)	.13	.20	-.03	.12
German words				(.30)	.24	.06	-.20
ESP					(.52)	.18	-.16
Ambiguous pictures						(.57)	-.10
Sex ( <i>r<sub>phi</sub></i> )							

Note.— $r = .14$  significant at .05;  $r = .18$  significant at .01.

used with a heterogeneous, noncollege population. However, when these demographic variables are controlled, a major portion of the reliable response set variance is eliminated.

### DISCUSSION

How large a correlation coefficient should be before it becomes meaningful is in many ways a value judgment. Many of the values in Table 1 are significantly different from zero at the .01 probability level. Edwards' (1950) distinction between statistical significance and practical significance is relevant here. In the search for sources of common variance a statistically significant correlation is of little practical value if a very small percentage of the variance of one scale is explained by its relationship with the other. The majority of coefficients in Table 1 indicate less than 7% of the variance of one scale is common to them both. Hence, for the most part, these data offer little evidence that a common response tendency was influencing the selection of response options across the six scales.

Other causes of low correlations should be considered in drawing the conclusion stated above. The role of low reliabilities of individual scales in shrinking the correlations is, of course, important. Yet, when these correlations are corrected for attenuation, the interpretations suggested by the data are not changed. Moreover, the finding of low reliability for the response set scores is meaningful in and of itself. Cronbach (1950) has suggested that the most simple and effective demonstration of a response set is derived from assigning a score for the particular response tendency (number of agreement responses, weighted by intensity, in this study) and computing the internal consistency reliability for this score. Following Cronbach's logic, one finds in the present study little evidence that an agreement response style was operating reliably in four of the six measures.

A second possibility is that the method of scoring these scales confounded the agreeing set with the response set to take extreme positions. However, Komorita<sup>4</sup> compared scores derived from seven-point agree-disagree scales with scores derived from dichotomizing the options. His calculations were made on the same data collected for this study, and he found correlations above .90 for every scale so examined.

Finally, since the three noncontent scales were presented as experiments, it is possible that the subjects were led to expect something near a 50-50 split in the "correct" responses to the stimuli, and therefore, restricted their tendency to agree with more than 50% of the items on these measures. To the extent that this is true, it would lower the variance of these scales. However, one would expect this effect equally across the three measures; while it might have lowered the correlations between them and the content scales, it may well have raised the correlations between the noncontent scales. Yet, it is not evident that any tendency was exerting a common influence on the noncontent scales in this study.

The major finding in Table 1 is the apparent importance of meaningful verbal content in the items as a source of common variance. Clearly the SAS, F Scale, and ARS form a factor grouping. None of the other scales correlate with these content measures to the extent present within this grouping. It is also notable that no prominent factor grouping appears among the noncontent tests. This suggests that where subjects are given a chance to respond in an acquiescent manner to different types of ambiguous stimuli, a very small portion of their responses are explainable by a common response tendency. In the correlations across degree of stimulus meaningfulness, the data again fail to support the assumption of a response set variable common to the tests. Cronbach (1950) cites evidence that

<sup>4</sup> Personal communication.

acquiescence scores are correlated across similar situations, i.e., personality inventory items. (Here the average correlation reported is .24.) He also presents evidence that response sets are not consistent over widely varying situations. These data were derived from judgments of physical stimuli as either "pleasant" or "unpleasant." It now appears that when stimuli vary on the content meaningfulness dimension, the tendency to acquiesce likewise fails to appear consistently.

The most parsimonious interpretation of the data in Table 1 is that there is no general trait of response acquiescence which appears reliably in different situations; it is, rather, dependent upon specific groups of items used to measure it. That is, a "pure response set" may be a straw man. This conclusion is clearly suggested by these data, and it receives support from other independent research.

In two recent papers (Jackson & Messick, in press; Messick & Jackson, 1961) emphasis has been placed on variations in the acquiescence-eliciting properties of different item forms. The investigations of these authors (Messick & Jackson, 1961) have lead them to assert that

the acquiescence-evoking properties of items are not . . . uniform over all scales, but that acquiescence is elicited differentially as a function, perhaps, of specific item content, of the clarity or ambiguity with which the content is stated, and in particular of the perceived desirability of the statement (p. 302).

Therefore, perhaps one should not expect high correlation across scales which vary as grossly in item form as the six measures used in this study. Of course, the corresponding conclusion is that one should not expect to find a general trait of acquiescence either.

Gray and Crisp (1961) have also reasoned that the use of a response set score as a personality trait requires that the score generalize further than a single set of items. They obtained acquiescence responses to a modification of the Perceptual Reaction Test (Berg, Hunt, & Barnes, 1949) and to a completely content free task presented as a measure of subliminal perception. The two sets of scores differed markedly from one another in distribution, in mean acquiescence, and correlated only  $-.01$ . Clearly, a common behavioral tendency such as response acquiescence could not explain the responding to these two unstructured, nonverbal measures. They concluded that

it is still parsimonious and appropriate to regard with suspicion the notion of a generalized free-float-

ing response set utterly divorced from content (p. 11).

The findings of the present study, congruent as they are with other very recent research, have important implications for the new approach to response set scores as traits based on central personality syndromes. It does not appear feasible to continue thinking in terms of a pure response set independent of specific item content. Hence, in attempting to relate acquiescence to personality traits, scores derived from a single instrument are not sufficient to constitute the measure of the response style.

#### SUMMARY

This study is the first in a series of investigations designed to evaluate the currently popular hypothesis that response styles are based on personality traits, and thereby have utility in personality assessment. Responses of 218 subjects on six different scales of acquiescence, varying in degree of meaningful verbal content, were intercorrelated. The results revealed that only those scales containing similar verbal content in the items were related. Apparently verbal content is quite important, whereas the amount of structure of the items is less important in determining agreement responses, than has been previously supposed.

The data are interpreted as suggesting that there is no general trait of response acquiescence independent of specific instruments used to measure it.

#### REFERENCES

- ADORNO, T. S., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, 53, 296-299.
- BERG, I. A., HUNT, W. A., & BARNES, E. H. *The Perceptual Reaction Test*. Evanston, Ill.: Berg, 1949.
- BERG, I. A., & RAPAPORT, G. M. Response bias in an unstructured questionnaire. *J. Psychol.*, 1954, 38, 475-481.
- COUCH, A., & KENISTON, K. Yeasayers and nay-sayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, 60, 151-174.
- CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, 6, 475-494.
- CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, 10, 3-31.



- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
- GRAY, C. W., & CRISP, H. E. The credibility of pure response set. Paper read at Southeastern Psychological Association, Gatlinburg, April 1961.
- JACKSON, D. N., & MESSICK, S. J. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243-252.
- JACKSON, D. N., & MESSICK, S. J. Acquiescence and desirability as response determinants on the MMPI. *Educ. psychol. Measmt.*, in press.
- McGEE, R. K. Response style as a personality variable: By what criterion? *Psychol. Bull.*, 1962, in press.
- MESSICK, S. J., & JACKSON, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychol. Bull.*, 1961, **58**, 299-304.
- NUNNALLY, J. C., & HUSEK, T. R. The phony language examination: An approach to the measurement of response bias. *Educ. psychol. Measmt.*, 1958, **18**, 275-282.

(Received February 9, 1961)

## EXPRESSIVE MOVEMENTS, WARMTH, AND VERBAL REINFORCEMENT<sup>1</sup>

MICHAEL M. REECE AND ROBERT N. WHITMAN

*Wayne State University*

Recent studies have applied the principle of reinforcement in an effort to influence verbalization. Krasner (1958) has reviewed numerous studies in which verbal behavior has been investigated in terms of instrumental learning. A wide variety of stimuli has been found to be effective as secondary reinforcers; e.g., gestures, smiling, nodding, and leaning forward. Verbalization has been shown to have secondary reinforcing properties. Experimenters frequently have used "mm-hmm," "good," and "fine" as a means of reinforcement. These words, and similar ones having the connotation of approval, can be interpreted by the listener as indicating the attitude of the therapist.

Most views of psychotherapy agree on the importance of the therapist's attitude toward the client and the client's interpretation of it. Rogers (1951) has stressed the importance of acceptance, permissiveness, and warmth on the part of the therapist. Fiedler (1953) investigated the influence of the therapist's feelings upon the client and concluded from the evidence he presented that the therapist seems to be "primarily responsible for the character of the therapeutic relationship" (p. 315). It also has been suggested that warmth and acceptance by the therapist can be considered in terms of cues indicative of intense, undivided attention (Krasner, 1955).

In order to be effective, the therapist's attitudes and feelings must be conveyed to the client. They may be communicated verbally and nonverbally. Inferences can be made from the content of the therapist's speech and from the expressive movements which are part of his behavior.

It may be considered that a therapeutic climate of warmth, with the connotation of permissiveness and acceptance, has a reward value for the client, hence a reinforcing effect for learning. The present study was designed to determine whether warmth has a reinforcing influence upon verbal behavior; and particularly, whether warmth combined with verbal reinforcement would produce a greater total reinforcing effect upon verbalization.

### METHOD

The experimental procedure consisted of a 15-minute, "free association" situation. Each subject was asked to say whatever words came to mind and was instructed to say each word individually, not to use phrases or to count. The responses were recorded on tape.

A factorial design was employed with 69 college students who were randomly assigned to four groups designated as warm-reinforced, warm-nonreinforced, cold-reinforced, and cold-nonreinforced.

"Warm" and "cold" were defined in terms of the experimenter's expressive movements, based on a previous study which had related approach and avoidance movements to judgments of warmth and coldness (Reece & Whitman, 1961).

When "warm," the experimenter leaned toward the subject, looked directly at him, smiled, and kept his hands still. When "cold," the experimenter leaned away from the subject, looked around the room rather than at him, did not smile, and drummed his fingers. The expressive movements were maintained continuously for the experimental period. An experienced judge periodically observed the experimenter through a one-way mirror to check the consistency of his movements defining "warm" and "cold."

With the designated groups, plural nouns were reinforced by the experimenter's saying "mm-hmm." Except for the instructions, there was no other verbalization by the experimenter.

Upon completion of the task, each subject completed a questionnaire giving his reactions to the task and to the experimenter, judging him as "warm" or "cold," and indicating the basis for the judgment.

### RESULTS

The tapes were replayed and both the frequency of all verbal responses and the frequency of plural noun responses were tabulated. The instructions were followed effectively so that no proper nouns or numbers were given. In addition, the subjects apparently provided their own "set" against repetition of words. Repeated words were scored since repetition rarely occurred.

In order to determine the reliability of this counting procedure, another scorer independently recorded the frequencies of the responses in a sample of 20, randomly selected, 3-minute periods of the tapes. These scores were correlated with those found by the experimenter for the same periods. The Pearson  $r$  for all verbal responses

<sup>1</sup> Based on an MA thesis by the junior author under the direction of the senior author.

TABLE 1  
MEAN TOTAL VERBAL RESPONSES

Group	N	M	SD
Warm reinforced	16	346.94	114.99
Warm-nonreinforced	15	310.67	117.19
Cold-reinforced	20	298.70	117.01
Cold-nonreinforced	18	298.33	84.93

was .99 and for plural noun responses was .98, indicating very high interscorer consistency.

It is seen from Table 1, which gives the means for all verbal responses, that the warm-reinforced group produced the greatest number of words, and that more verbalization occurred in both warm groups than in the cold groups.

The analysis of variance, shown in Table 2, indicated a significant interaction between verbal reinforcement and expressive movements, and showed that the expressive movements variable also was highly significant.

Marked differences were found with the specifically reinforced plural nouns, as shown in Table 3. The superiority of the reinforced groups was found to be definitely significant by the analysis of variance method (McNemar, 1955) as indicated in Table 4.

Evaluation of the extent to which the subjects' judgments of "warm" and "cold" agreed with the expressive movements defining these conditions yielded a chi square of 20.41 ( $n = 3$ ), significant at less than the .01 level.

When the interview data were examined, it was found that most of the subjects were aware of the expressive movements and utilized them as a basis for their judgments. However, some of the subjects, especially in the cold-reinforced group, appeared to be more responsive to the experimenter's reinforcing verbalization than to the expressive movements.

#### DISCUSSION

It is evident from the results of this study that the "climate," as defined by expressive

TABLE 2  
ANALYSIS OF VARIANCE FOR TOTAL  
VERBAL RESPONSES

Source	df	MS <sup>a</sup>	F
Reinforcement (R)	1	4,694.26	.86
Expressive movements (EM)	1	16,435.62	13.01***
Interaction R × EM	1	5,492.04	4.35*
Within	65	1,263.37	

<sup>a</sup> Corrected for unequal N by Snedecor's (1956) method.

\* Significant at .05 level.

\*\*\* Significant at .001 level.

TABLE 3  
PLURAL NOUN RESPONSES

Group	N	M	SD
Warm-reinforced	16	82.44	58.44
Warm-nonreinforced	15	40.00	24.68
Cold-reinforced	20	64.85	40.13
Cold-nonreinforced	18	37.17	30.23

movements, was an effective variable. The interaction of warm expressive movements with verbal reinforcement produced the greatest amount of verbalization when all words were considered. Verbal reinforcement alone was not a significant influence upon the total number of spoken words. On the other hand, the expressive movements did prove to have a significant effect.

Since the verbal reinforcement awaited the subjects' responses, the number of reinforcements was not the same for the reinforced groups. Reinforcement therefore would not be expected to produce a uniform result for these groups.

Many studies have shown that verbal reinforcement of a specific class of verbal responses will increase the frequency of those responses. The present findings support and reaffirm this conclusion. The reinforcement of saying "mm-hmm" when plural nouns were voiced was sufficient to produce a significant increase in this response.

Since the warmth or coldness of the experimenter was present throughout a particular condition and was not restricted to plural nouns or specifically associated only with them, it is not surprising that the expressive movements had no significant effect upon this class of response.

It was clearly demonstrated, also, that the expressive movements were not mere manipulations by the experimenter but successfully conveyed the impression of the experimenter's attitude. Leaning toward the subject, smiling, and looking directly at him enabled the subject to judge the experimenter as warm. Conversely, looking away from the subject, leaning away from him, not smiling, and intermittently drum-

TABLE 4  
ANALYSIS OF VARIANCE FOR PLURAL NOUNS

Source	df	MS <sup>a</sup>	F
Reinforcement (R)	1	20,269.14	11.72**
Expressive movements (EM)	1	1,881.24	1.09
Interaction R × EM	1	928.74	.54
Within	65	1,730.15	

<sup>a</sup> Corrected for unequal N by Snedecor's (1956) method.

\*\*  $p < .01$ .



ming the fingers on the table impressed the subject as coldness on the part of the experimenter.

#### SUMMARY

In order to determine the effects of the experimenter's warmth and coldness upon verbal conditioning in a free association task, 69 subjects were randomly assigned to four experimental groups in a factorial design.

Warmth and coldness were defined in terms of expressive movements of posture, glance, facial expression, and finger activity.

Analysis of the results showed that the total number of words was significantly affected by the expressive movements and verbal reinforcement, but was not significantly influenced by verbal reinforcement alone.

The number of plural nouns was significantly increased by verbal reinforcement but was not significantly affected by the expressive move-

ments or by the interaction of movements and reinforcement.

#### REFERENCES

- FIEDLER, F. E. Quantitative studies on the role of the therapists' feelings toward their patients. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald, 1953. Pp. 296-325.
- KRASNER, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, 1, 19-25.
- KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, 55, 148-170.
- MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.
- REECE, M. M., & WHITMAN, R. N. Warmth and expressive movements. *Psychol. Rep.*, 1961, 8, 76.
- ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.
- SNEDECOR, G. W. *Statistical methods*. Ames: Iowa State Coll. Press, 1956.

(Received January 30, 1961)

## OVEREVALUATION OF OWN GROUP'S PRODUCT IN INTERGROUP COMPETITION<sup>1</sup>

ROBERT R. BLAKE AND JANE SRYGLEY MOUTON

*University of Texas*

Any social system which embraces values such as excellence, efficiency, or adequacy as standards for performance develops normative criteria by which the achievements of individual, groups, or organizations can be evaluated. Depending on results from applying such judgments, the evaluated unit, whether individual, group, or organization, may be accepted or rewarded, rejected or punished, either completely or in varying degrees. The judgments leading to acceptance or rejection may be self-imposed or made by others. Grant that, under such conditions, a fundamental human motive is to increase the probability of acceptance or to avoid the risk of rejection. Then, acceptance generates security and reduces anxiety; rejection increases anxiety and reduces security.

When absolute normative criteria of excellence, efficiency, or adequacy are unavailable, as they frequently are, judgments only can be made in relative terms. With criteria which are relative rather than absolute and with motivation in the direction of increasing security through being excellent, efficient, or adequate then distortions in perception and judgment are possible. Such distortions, it is predicted, are in the direction of evaluating one's own performance above that of a comparison individual or individuals. When the object of evaluation is a group product and one values his group membership, the prediction is that distortions of judgments occur in the same direction as above; that is, members evaluate their own group product as being "better" than the product of a comparison group or groups.

The present paper describes an experiment designed to evaluate the accuracy of this prediction under conditions where group products are being compared. The specific hypothesis tested is that group members elevate the evaluation of their own solution to a problem above their evaluation of the product from a comparison group.

### PROCEDURE

#### *Setting and Subjects*

The experiment was conducted in 17 different laboratory training programs, representing 48 groups and totaling 410 subjects. Competition took place

between eight pairs, four trios, and five quartets of groups, with groups composed of 8-10 members each.

Participants in 32 of the 48 groups were adults who ranged in age from 25 to 55 and who were engaged in some aspect of administrative work in connection with industrial, medical, and research organizations. The remaining 16 groups were composed of junior and senior college students, both male and female, enrolled in social psychology. The sequence of activities was the same for adult and college subjects. All groups were matched to equalize membership in terms of educational and occupational level for adult groups and in class level and sex for college groups.

#### *Sequence of Interactions*

The arrangements for creating intergroup competition were:

*Development of autonomous groups.* Ten to 12 hours were devoted to autonomous ingroup activity in which individuals met together to study decision making in groups through developing, and then evaluating, their own ingroup structures.

*Creating intergroup competition.* After the initial period of ingroup development, an experimental arrangement was created such that groups found themselves in competition with each other. Each group had 3 hours to formulate its own approach to the solution of an assigned problem. The problem, to be solved in the form of a two-page, double-spaced, typewritten memorandum, concerned some issue within an organization or intergroup relations with which all members of the contending groups were equally familiar. Group members routinely accepted the proposition that their performance on the assigned task provided an indication of their adequacy as a problem solving unit. A winner and a loser were to be determined. The possibility of resolving differences between solutions by compromise was unavailable.

*Comparison of group positions.* Solutions were then dittoed and distributed to all competing group members for purposes of comparison and clarification of differences. Members judged the merits of the contending solutions in private. In some of the competitions, the rating scale read, "How do you feel about the solutions?" with judgments ranging from 9, "best possible," to 1, "worst possible." In others, the scale was, "Rate the solutions in terms

<sup>1</sup> These studies were supported by Grant M-2477, Behavior of Group Representatives under Pressure, National Institutes of Health.

of overall adequacy," from 9, "completely adequate," to 1, "completely inadequate." Findings were unrelated to the scale used.

### RESULTS

Results were analyzed in several different ways. One was a comparison of judgments by individuals of own and other group products. Since private individual judgments were made within the context of group interaction the group has been taken as the unit of analysis and group averages of judgments of own and other's products have been compared also. On a group basis, there were 48 possible comparisons between the average group evaluation of own position relative to the proposal created by another. When there were more than two groups in competition with one another, only the highest evaluation of the contenders' products was used for purposes of analysis. For 46 of the comparisons, the product of one's own group was rated higher than the competitor's and in two they were rated equal. In no instance was the average group rating of own solution rated lower than an opponent's solution. The  $\chi^2$  of 40.33 ( $df = 1$ ) is significant beyond the .001 level. On an individual-by-individual basis, there were 410 possible comparisons between judgments of one's own group product and that proposed by a contending group. Own group product was rated higher than the comparison group in 378 of the instances, own product was rated as equal in quality 19 times, and in only 13 instances did a group member rate his own group product as inferior to that created by another group. The  $\chi^2$  of 291.99 ( $df = 1$ ) also is significant beyond the .001 level.

The average judgment by the 410 subjects of own group product was 7.5 and of the product of the comparison group, 5.6. Where there were three or four groups judging together the highest judgment of the other group's product was used

for comparing the mean ratings. The  $t$  for the differences is 23.84 ( $p < .001$ ,  $df = 409$ ). The comparable averages computed on a group basis were 7.6 for own group product, and 5.5 for the product created by the comparison group, with a  $t$  of 3.21 ( $p < .01$ ,  $df = 47$ ). The results demonstrate that group members uniformly judge their own solutions as being better in quality than they do the solutions of comparison groups.

### DISCUSSION AND CONCLUSIONS

Four hundred and ten persons, participating as members of 48 groups in human relations training programs, engaged in a parallel sequence of ingroup and intergroup activities. Groups were placed in competition in pairs, trios, and quartets of groups, with the outcome determined by the quality of the solutions created by the groups. Quality judgments were made by group members on a private, or personal, basis for the adequacy of their own and of the other group's positions. They resulted in one's own product being rated higher than that created by a competitor's.

The results confirm the prediction that group members evaluate their own group product above the judgments they accord to the proposal from a comparison group.

The overevaluation of one's own group product relative to a comparison group can be interpreted in several ways: (a) as due to perceptual distortion stemming from group identification and needs in a situation where personal adequacy via group adequacy serves as a criterion of acceptance or rejection under win-lose conditions; (b) as stemming from distortions in reporting evaluations for the instrumental purpose of "winning," or because of greater familiarity with the rationale, reasons, and premises of one's own group's solution.

(Received February 2, 1961)



## TIME ORIENTATION IN MALE AND FEMALE JUVENILE DELINQUENTS<sup>1</sup>

ANTHONY DAVIDS, CATHERINE KIDDER, AND MELVYN REICH

*Brown University and Emma Pendleton Bradley Hospital*

A recent review by Wallace and Rabin (1960) indicates a long history of psychological interest in temporal experience as well as an increasing amount of research activity in this area of investigation. In this journal, Barndt and Johnson (1955) reported that delinquent boys are more present oriented than are nondelinquent boys. In their study, the delinquent and nondelinquent subjects were well-matched on variables such as age, IQ, school achievement, and socioeconomic status. Utilizing a story completion technique, they found that the delinquent boys produced stories with significantly shorter time spans than did the boys in the control group.

A later investigation, reported in this journal by Davids and Parenti (1958), found no differences between emotionally disturbed and normal 11-year-old boys with respect to time orientation. However, both the disturbed and normal subjects in this study were found to be more present oriented than the normal adolescents in the Barndt and Johnson (1955) investigation, but were not different from the adolescent delinquent boys in this respect. Since the Davids and Parenti study utilized the identical story completion technique and scoring procedures from the Barndt and Johnson study, it was possible to make these direct comparisons between findings obtained from different age subjects who were studied in independent investigations conducted in quite different institutional settings. However, since our earlier extension of the Barndt and Johnson research led to findings that appeared somewhat paradoxical, it was decided to attempt an exact replication of their study, using adolescent subjects institutionalized for delinquency, rather than younger children in a residential psychiatric treatment center.

### METHOD

**Subjects.** There were two groups of subjects in this investigation. One group consisted of 24 boys

<sup>1</sup> This research was supported by Small Grant M-4327 (A) from the National Institute of Mental Health, United States Public Health Service. We wish to express our appreciation to Anthony Trivisono, who was Superintendent of the Rhode Island Training School for boys, and Donald D. MacDougald, who is Superintendent of the Rhode Island Training School for girls, for facilitating this study.

who were institutionalized in the Rhode Island Training School for boys, and the other group consisted of 30 girls institutionalized in the Rhode Island Training School for girls. The mean age in both groups was between 15 and 16 years, and all of these adolescents had been sentenced by the court for delinquent activities.

**Procedure.** The procedure followed was previously employed and described by Barndt and Johnson (1955). The subject was given the following instructions:

I want to see what kind of a story you can tell. I'll start a story and then let you finish it any way you want to. You can make it any kind of a story you wish. Do you understand? I'll start it now.

The story began as follows:

About three o'clock one bright sunny afternoon in May, two boys [or two girls in the procedure with the female subjects] were walking along near the edge of town . . .

After reading the beginning of this story to the subject, the experimenter said,

Now you start there and finish the story any way you want to.

The experimenter then recorded the subject's story verbatim. After finishing the story, the subject was asked, "What time was this?" or "How long did the story take to happen?" The scoring system described by Barndt and Johnson was used. The scoring categories were as follows: 1—under 1 hour; 2—1 hour to under 5 hours; 3—5 hours to under 12 hours; 4—12 hours to under 1 week; 5—1 week to under 3 months; 6—3 months or more. The amount of time covered in the stories could be scored very accurately, and agreement between raters was unity.

### RESULTS AND DISCUSSION

The results obtained in this study are shown in Table 1. The findings indicate identical mean time orientation scores in the male and female groups of adolescent delinquents. Table 1 also contains findings obtained in independent studies published previously. It is interesting to note that the time orientation scores of the present groups of adolescent delinquents are very similar to scores obtained by Davids and Parenti (1958) in a study of 11-year-old boys who were institutionalized for severe emotional disturbances and 11-year-old normal boys who were attending a public school. Moreover, the time orientation

TABLE 1  
COMPARISON OF PRESENT DELINQUENT GROUPS AND GROUPS STUDIED PREVIOUSLY

Measure	Groups					
	Present delinquent girls ( <i>N</i> = 30)	Present delinquent boys ( <i>N</i> = 24)	D-P disturbed boys ( <i>N</i> = 30)	D-P normal boys ( <i>N</i> = 22)	B-J delinquent boys ( <i>N</i> = 26)	B-J normal boys ( <i>N</i> = 26)
Mean story score	2.7	2.7	2.9	2.9	2.8	3.6
Variance	2.4	1.7	2.7	1.4	2.0	1.7
Mean age	15.5	15.2	10.10	10.10	16.8	16.9

Note.—D-P refer to findings from Davids and Parenti (1958). B-J refer to findings from Barndt and Johnson (1955).

scores of the present groups of delinquents do not differ significantly from the scores obtained in the Barndt and Johnson (1955) study of institutionalized adolescent delinquents. Not only are the mean scores very similar, but according to the *F* test for homogeneity of variance, there are no significant differences among the variance scores. It is noteworthy that the same assessment procedure utilized several years ago with delinquents in the Midwest yields almost identical findings from delinquents in New England.

Comparison of the present groups of delinquents with Barndt and Johnson's (1955) group of normal adolescent boys showed significant differences in keeping with their findings. Our group of delinquent boys was found to be significantly more present oriented than their normal boys ( $t = 2.43$ ,  $p = .05$ ). The difference between our delinquent girls and their normal boys was even more significant ( $t = 6.04$ ,  $p = .001$ ). Thus, on the basis of these studies, it appears that normal adolescents tend to be more future oriented than do male and female adolescents who engage in serious delinquent activities. It should be emphasized, however, that without a more comprehensive sampling design, these results should be viewed as suggestive rather than definitive.

Our previous findings revealed no significant differences in time orientation in normal and emotionally disturbed 11-year-old children. In fact, these earlier findings suggested that present orientation was correlated with stable friendships and future orientation was correlated with personality characteristics of pessimism and resentment. It was speculated that for children in the latency stage it may be "normal" to be concerned mainly about the present and that concern for the future might be an index of personal dissatisfaction and social instability. However, adequate understanding of these perplexing matters awaits

an exhaustive developmental study of normal and maladjusted children at numerous stages of maturation and development. It may well be that what is "normal" or socially desirable at one stage of ego development is pathological at some earlier or later stage in the developmental process.

#### SUMMARY

This paper represents an attempt to relate findings obtained from three independent studies. Using a story completion technique to assess time orientation in imaginal processes, no difference was found between adolescent delinquent boys and adolescent delinquent girls. The measures of time orientation obtained in this study were practically identical with findings obtained in a previous study of adolescent delinquent boys. Moreover, time orientation revealed in the fantasies of these three groups of delinquents was very similar to time orientation evidenced by much younger emotionally disturbed boys and normal boys. Although there were no significant differences among these five groups, all of them were found to be significantly more present oriented than were a group of adolescent normal boys. Thus, the major finding was a significant difference between time orientation in delinquent and normal children.

#### REFERENCES

- BARNDT, R. J., & JOHNSON, D. M. Time orientation in delinquents. *J. abnorm. soc. Psychol.*, 1955, 51, 343-345.  
 DAVIDS, A., & PARENTI, ANITA N. Time orientation and interpersonal relations of emotionally disturbed and normal children. *J. abnorm. soc. Psychol.*, 1958, 57, 299-305.  
 WALLACE, M., & RABIN, A. I. Temporal experience. *Psychol. Bull.*, 1960, 57, 213-236.

(Received March 1, 1961)

# Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

## CONSULTING EDITORS

JOSEPH B. ADELSON  
*University of Michigan*

GORDON W. ALLPORT  
*Harvard University*

NORMAN H. ANDERSON  
*University of California, Los Angeles*

LEONARD BERKOWITZ  
*University of Wisconsin*

JACK W. BREHM  
*Duke University*

URIE BRONFENBRENNER  
*Cornell University*

DONALD R. BROWN  
*Bryn Mawr College*

DONALD T. CAMPBELL  
*Northwestern University*

LAUNOR F. CARTER  
*System Development Corporation*

DWIGHT W. CHAPMAN, JR.  
*Vassar College*

MORTON DEUTSCH  
*Bell Telephone Laboratories*

SEYMOUR FESHBACH  
*University of Pennsylvania*

DONALD W. FISKE  
*University of Chicago*

NATHANIEL L. GAGE  
*University of Illinois*

NICHOLAS HOBBS  
*Peabody College for Teachers*

MURRAY HORWITZ  
*New York University*

IRVING L. JANIS  
*Yale University*

HAROLD H. KELLEY  
*University of California, Los Angeles*

HERBERT C. KELMAN  
*Harvard University*

DANIEL J. LEVINSON  
*Harvard Medical School*

DAVID C. MCCLELLAND  
*Harvard University*

ROBERT B. MACLEOD  
*Cornell University*

HAROLD L. RAUSH  
*National Institute of Mental Health*

NEVITT SANFORD  
*Stanford University*

IRVING SARNOFF  
*Western Reserve University*

WILLIAM A. SCOTT  
*University of Colorado*

JOSEPH VEROFF  
*University of Michigan*

JUSTIN L. WEISS  
*Massachusetts Mental Health Center*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.



The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is \$10.00, or \$20.00 annually (foreign, \$20.50). Single issues are \$2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ  
University of Michigan  
Box 1248  
Ann Arbor, Michigan

---

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR  
Acting Production Manager

JUDY SMITH  
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 Sixteenth Street, N.W.  
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
Prince and Lemon Sts., Lancaster, Pa.  
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Lancaster, Pa., and at additional mailing offices

Copyright © 1962 by the American Psychological Association, Inc.

## PERSISTENCE OF THE RESISTANCE TO PERSUASION INDUCED BY VARIOUS TYPES OF PRIOR BELIEF DEFENSES<sup>1</sup>

WILLIAM J. MCGUIRE

*Department of Social Psychology, Columbia University*

A number of previous studies have tested the relative efficacy of various types of prior defenses in making a person's belief resistant to change when he is later confronted with massive counterarguments against the belief. None of these previous studies were designed to measure the effect on resistance of varying the time interval between the defense and the attack. By systematically varying this interval, the present experiment investigates the relative persistence of the immunity of persuasion conferred by the different types of prior defenses. It is of some interest to know for each type of defense the rate at which its conferred immunity decays over time. Of even greater theoretical interest are comparisons among the decay rates for the different types of defense.

The predictions regarding these differential decay rates derive from the same postulates as gave rise to the earlier predictions, tested and confirmed in previous experiments, regarding the relative immunizing effectiveness of various defenses *without* regard to the time interval between defense and attack. Hence, it is useful to mention several of the relevant previous findings and their theoretical bases. The previous studies, and the present one as well, used cultural truisms as the beliefs being defended and attacked—for example, the belief that "We should brush our teeth after every meal if at all possible." It had been postulated that there is little belief-dissonant information available regarding such cultural truisms in the person's normal ideological environment. This unavailability combined with the characteristic tendency to avoid even such belief-dissonant material as is available, would have left the person underestimating the vulnerability of his belief and,

hence, unmotivated to acquire bolstering material and unprepared to deal with strong counterarguments when he is forced to expose himself to them.

From this theoretical analysis follow several of the previously confirmed hypotheses regarding immunization against persuasion which are relevant to the hypotheses being tested in the present experiment. One of these previous findings (McGuire & Papageorgis, 1961) is that prior refutational defenses are superior to prior supportive defenses in making cultural truisms resistant to subsequent persuasion. Refutational defenses involve mention and refutation of possible counterarguments against the belief, while ignoring arguments positively supporting the belief. Supportive defenses do mention and elaborate arguments positively supporting the belief, while ignoring possible counterarguments against it. This superiority of the refutational defense would follow from the above theoretical assumptions, since the refutational defense contains a threatening element—mention of the counterarguments to whose existence the subject has probably given little, if any thought—which stimulates him to bolster his belief. The supportive defense of the truism, on the other hand, seems to labor the obvious, so that the subject is little motivated to assimilate the positive arguments and is left, if anything, even less motivated to bolster further the belief he regards as obvious.

It was also demonstrated (Papageorgis & McGuire, 1961) that the refutational defense confers resistance to subsequent attacks even by novel counterarguments, different from those explicitly refuted in the defense. This conferral of generalized immunity by the refutational defense also follows from the theoretical assumptions. The immunizing efficacy derives not only from weakening the credibility of the specific counterarguments

<sup>1</sup> This study was supported, in part, by a grant from the National Science Foundation, Division of Social Sciences.

refuted, but also from the threat induced stimulation to bolster one's defense. Hence, the refutational defense increases resistance even to attacks by counterarguments other than those refuted.

The foregoing theoretical interpretation of the previous findings gives rise to three predictions regarding the temporal persistence of the resistance conferred by the different types of prior defense. First, it is hypothesized that the supportive defense will not only be initially inferior to the refutational in the amount of resistance it confers, but in addition that such resistance, as it does confer, will decay more rapidly than that conferred by the refutational defense. This prediction follows from the above interpretation that the immunizing efficacy of the supportive defense derives solely from the acquaintance with the positive arguments which it contains and which tend to be forgotten over time; the efficacy of the refutational defenses, on the other hand, derives in part from the threat induced motivation to bolster one's defenses. Since for some time the subject will continue to act on the motivation, the forgetting of the refutational material will be partly offset by this continued acquisition of bolstering material.

The second hypothesis is that the temporal decay of conferred immunity occurs more rapidly against attacks by the same counterarguments as had been explicitly refuted than against attacks by novel counterarguments. The theoretical basis for this prediction is quite similar to that yielding the first hypothesis. The immunity to attacks by the very counterarguments refuted derives from both recall of the specific refutations, which decays over time, and the amount of bolstering material the subject has acquired on the basis of his induced motivation, which increases over time. The immunity to attacks by novel counterarguments derives solely from the latter mechanism. Hence, conferred resistance to novel counterarguments should tend to catch up over time with resistance to the very counterarguments refuted.

The third hypothesis, which follows as a corollary of the second, is that the refutational defense has a delayed-action effect in conferring resistance to attacks by novel

counterarguments. The refutations per se should not confer any resistance in this case, at least, not in so far as the counterarguments used in the attack are indeed novel. Hence, any resistance conferred derives from the second mechanism, the motivation to bolster one's belief induced by exposure to the threatening counterarguments during the defense. Acting on this motivation requires time, particularly in the monolithic ideological environment that tends to surround cultural truisms. Hence, the resistance to attack by novel arguments will continue to grow for sometime after the threatening pre-exposure. As time passes, the induced motivation will, of course, itself decay so that the total time function will be nonmonotonic. But for a time at least the conferred immunity will grow.

To test these temporal-trend predictions adequately it is important that we have some idea of the time parameters involved. It was in part to explore these parameters that the time interval between defense and attack was deliberately varied, from experiment to experiment, in the previous studies in this series. For example, in McGuire (1961a), the attack came immediately after the defense; in McGuire and Papageorgis (1961), 2 days intervened; and in Papageorgis and McGuire (1961), the interval was 1 week. Hence, it is possible to make a crude test of the three temporal-trend hypotheses by cross-experimental comparisons. The results based on such cross-experimental comparisons are depicted in Figure 1a and can be seen to be in accord with each of the three hypotheses. This confirmation of the predictions cannot be regarded as definitive since extraneous conditions varied somewhat from experiment to experiment. For example, the issues, defensive and attacking messages, and types of subjects all differed somewhat among the experiments. The confirmations are sufficiently clear, however, that we were encouraged to vary systematically the time intervals within the present experimental design over comparable magnitudes.

#### METHOD

*Procedure.* The study was represented to the subjects as an investigation of personality correlates of



verbal skills, a deception that was bolstered by several tasks the subjects were called upon to perform during the experimental sessions. Each of the 160 subjects took part in two experimental sessions. During the first they received 600-word mimeographed messages defending their initial beliefs on medical truisms such as "Everyone should visit his doctor at least once a year for a routine physical check up." The subject was told that he would be scored on his ability to analyze such technical passages and he was given 4 minutes to read and, in each paragraph, underline the shortest clause that summarized the main point being made in the paragraph. This underlining task was introduced to encourage careful reading and to disguise the persuasive purpose of the messages. He was then given various personality tests, not relevant to the hypotheses under discussion, to disguise further the persuasive intent of the study.

The second session came either 2 days (for 80 subjects) or 7 days (for the other 80) later. In the second session, each subject received further defensive messages on additional medical truisms and then, within the same booklet, additional messages attacking the previously defended truisms and, in control conditions described under Design, previously undefended truisms as well. As in the first session, the subject was given 4 minutes to read and underline the crucial clauses in each of these defensive and attacking messages. Another personality questionnaire was then administered and then the subject was asked to fill out an opinion questionnaire indicating his own beliefs on the medical issues dealt with in the messages, on the pretext that we wished to check on whether the subject's personal opinions on the topics discussed in the passages affected his ability to read these passages analytically. The subjects then filled out a questionnaire designed to ascertain the extent to which the desired experimental conditions obtained,<sup>2</sup> after which the true

<sup>2</sup> This Critique of the Experiment final questionnaire was designed to measure the adequacy of the time allowances, how much the subject had heard of the experiment in advance and whether he suspected its persuasive intent. About 20% of the subjects complained that some section of the test had been given either a too long or a too short time allowance; more than half the complaints were that the time allowed for the noncrucial personality test was too short. The subjects were indeed rushed through this section to keep the session down to 50 minutes. A surprising number admitted having heard something about the experiment in advance, despite our request to the subjects that they refrain from discussing it with anyone until the end of the experimental period. Hearing that the test involved a reading comprehension test or that it dealt with medical topics was admitted by 31 out of the 160 subjects. In addition, 4 heard that one's opinions were measured. When called upon to suggest what else—besides verbal skills—the experiment could have measured, only 5 suggested any purpose having to do with opinion change, persuasion, or propaganda.

nature of the experiment and the nature of and reasons for the deceptions were explained to the subject.

*Defensive and attacking messages.* Two types of defensive messages were employed. The "supportive" defense had an introductory paragraph mentioning that the truism in question was obviously valid but that it was wise to consider some of the reasons why it was indeed valid. Two arguments in support of the belief were then mentioned. There followed two paragraphs each developing in a calm, factual way one of the two supportive arguments. These supportive messages avoided any mention of possible counterarguments against the truism.

The "refutational" defenses began with a similar introductory paragraph mentioning that the truism was obviously valid but that, since occasionally one heard misguided counterarguments attacking it, it was wise to consider some of these counterarguments and show wherein they erred. Two counterarguments against the truism were then mentioned. The following two paragraphs each refuted in a calm, factual way one of these counterarguments. These refutational messages avoided mention of arguments directly supporting the truism—they merely refuted counterarguments against it.

The attacking messages were similar in format to the defensive, each being about 600 words in length and divided into three paragraphs. The introductory paragraph stated that most laymen would be surprised to learn that advanced medical and scientific work was beginning to cast some doubt on the validity of the truism in question and, hence, it would be wise to ponder some of these recently discovered counterarguments against the belief, two of which were then mentioned. Each of the following two paragraphs expounded in a calm, factual manner the validity of one of these counterarguments. When the attacking message followed a refutation defense, one of two alternatives was used. Half of the subjects received attacks employing the very counterarguments refuted; the other half of the subjects received attacks employing novel counterarguments, different from those previously refuted.

Since the experimental design called for each subject's serving in four different defensive conditions, it was necessary to prepare supportive defense, refutational defense, and attacking messages on four different truisms. Furthermore, since half the refutational defenses had to be followed by attacks employing the same counterarguments as refuted and half by novel counterarguments, it was necessary to prepare two alternative forms of the refutational defense message and of the attacking messages on each issue, each dealing with a different pair of counterarguments. For symmetry of design, we prepared alternative forms of the supportive defense on each issue, each form using a different pair of supportive arguments. Hence, 24 messages in all were employed in the present study, 6 on each of four issues.<sup>3</sup>

<sup>3</sup> All 24 of the defensive and attacking messages used in this study have been deposited with the American Documentation Institute. Order Document

*Opinion questionnaire.* Beliefs on the four issues were measured by a 17-item opinion questionnaire. Each item consisted of an assertion on one of the issues (e.g., "Everyone should brush his teeth after every meal if at all possible.") followed by a graphic scale containing 15 numbered categories with Definitely False at one end and Definitely True at the other. The subject was told to make an "X" in whichever of the categories best indicated his own agreement with the statement. There were four items on each issue.<sup>4</sup> The scores cited in the Results section and in Table 1 are based on the mean of the four items on each issue, with the possible range going from 1.00, for complete rejection of the truism, to 15.00 for complete agreement therewith. One of the 17 items was a repeat to serve as a reliability check. The two responses to this repeated item yielded an intrasubject correlation of .82.

*Experimental design.* The design included four blocks of subjects. The subjects in the first block received refutational defenses on all four issues. They received such defenses on two issues in a first session 2 days previous to the attack; and the defenses on the other two issues at the second session immediately before the attack. In each session the defense on one issue involved refutations of the same counterarguments as would be used in the attack, and on the other issue, the refutation of alternative counterarguments to those that would be used in the attack. The subjects in the second block received the same treatments as those in the first block, except that for them the first session preceded the attacks by 1 week rather than 2 days.

The subjects in the third block received defenses on only two issues. Both of these were supportive defenses, one coming in the first session 2 days before the attack and one in the second session just before the attack. As regards the other two undefended issues, one was attacked in the second session to ascertain the impact of the attacks in the absence of any prior defense and one was not attacked to obtain an estimate of the initial belief levels in the absence of both defense and attack. The subjects in the fourth block received the same treatments as those in the third, except that for them the first session preceded the attacks by 1 week rather than 2 days.

Since there were four issues and two alternative sets of materials on each issue, eight subconditions were necessary in each block to allow the materials to be systematically rotated around the four treatment conditions. Five subjects served in each of these

eight "materials" subconditions, so that 40 subjects served in each of the four blocks.<sup>5</sup>

The purpose of this rather complex design was to allow more sensitive tests of the theoretically relevant and likely-to-be-small treatment effects. Thus, all comparisons between refutational defenses, for example, those predicted in the second and third hypotheses, involve intrasubject analyses. Likewise, comparisons between the supportive defense (which usually has the smallest effect—see for example, McGuire & Papageorgis, 1961) and no defense conditions also involve sensitive intrasubject analyses. It is true that comparisons between the refutation defenses on the one hand and the supportive defense, or the no defense control conditions involve across-subject comparisons, but such variations have been demonstrated in previous studies to produce sizable differentials.

The complexity of the design did necessitate the computation of several different error terms to evaluate the differential effects presented in the Results section below. In general, the error terms are based on the residual variance in the conditions being compared. The individual differences variance was removed when the comparison involved repeated measures on the same subject, for example, the effects of refutation of same vs. alternative counterarguments; or an interaction effect between type of defense and time of attack; or the effect of a supportive defense vs. no defense. When the comparison was between a refutational defense and a supportive or a no defense condition, the large between-subject residual variance, including individual differences among the subjects, was used as the error term.

*Subjects.* All 160 subjects were selected from a pool of students enrolled in the introductory psychology course at the University of Illinois. Among those who indicated at the beginning of the semester that they were regularly available on the days and hours chosen for running the experiment, the selection was random. About 75% of the 220 subjects requested to appear actually participated in the experiment and the data reported below are based on the first 160 of those who appeared for both sessions. The majority were sophomores and about 55% were females.

## RESULTS AND DISCUSSION

*General effects.* The two control conditions set the probable limits within which the differential immunization effects can take place. The mean belief level in the neither-attack-nor-defense control condition is 11.74 on the 15-point scale, and can be taken as an estimate of the initial belief level on the four truisms. Actually the estimate is probably conservative, since it is based on the indi-

No. 7058 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$2.25 for microfilm or \$5.00 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>4</sup> This opinion questionnaire has been deposited with ADI and can be obtained by writing for the document mentioned in Footnote 3.

<sup>5</sup> The design of the present study is described in detail in Table A of the ADI document mentioned in Footnote 3.



TABLE 1  
PERSISTENCE OF THE RESISTANCE TO PERSUASION CONFERRED BY  
THREE TYPES OF PRIOR BELIEF DEFENSE

Interval between defense and attack	Type of defense which preceded the attack			Attack without prior defense	Neither attack nor defense
	Supportive arguments	Refutation of counterarguments used in the attacks	Refutation of alternative counterarguments		
Immediate	9.71 (80) <sup>a</sup>	11.36 (80)	10.41 (80)	8.49 (80)	11.74 (80)
Two days	8.51 (40)	11.08 (40)	11.45 (40)		
Seven days	8.82 (40)	9.49 (40)	9.68 (40)		

Note.—Scores in the cells are final belief levels on the Truisms as measured on a 15-point scale.  
<sup>a</sup> Numbers in parentheses indicate the number of individual scores on which the cell mean is based.

cated belief on an issue unmentioned in the messages, but taken after the receipt of messages strongly attacking three other truisms and, hence, may reflect a general wariness on the part of the subject (see McGuire & Papageorgis, 1961, and Papageorgis & McGuire, 1961, for data on the accuracy with which such postcommunication beliefs on control, unmentioned issues estimate the initial level of the beliefs). The mean belief level in the other control condition (attack-only) is 8.49, indicating that in the absence of any defense, the attacks were effective in reducing the beliefs 3.25 points on the 15-point scale ( $p < .001$ ). The overall belief level in all three defense-and-attack conditions at all three time intervals is 10.17, which is almost exactly midway between the neither-attack-nor-defense and the attack-only means and significantly ( $p < .01$ ) different from either. Furthermore, the means in all nine defense-and-attack conditions (see Table 1) do lie between the two means of the neither-attack-nor-defense and attack-only conditions.

*Relative persistence after supportive and refutational defenses.* The supportive defenses conferred less resistance to the attack than did the refutational defenses regardless of whether the attack came immediately, 2 days, or 1 week after the attack. When the attack followed immediately, the superiority of the combined refutational defense conditions to the supportive was significant at the .01 level, but this superiority was primarily due to the conditions in which the very counterarguments used in the attack were refuted: the

superiority over the supportive reached only the .20 level of significance when counterarguments alternative to those used in the immediate attack were refuted.

Where 2 days intervened between the attack and defense, this superiority of the combined refutational to the supportive defense became even more pronounced ( $p < .001$ ). Whereas the immediate resistance conferred by the supportive defense had decayed ( $p < .05$ ) almost completely after the 2-day interval, that conferred by the refutational defense actually showed a trivial gain from the immediate to the 2-day interval. As can be seen in Table 2, this gain yielded an  $F$  of only 1.05. This interaction effect between the supportive vs. refutation type-of-defense variable and the immediate vs. 2-day interval variable is significant at the .01 level and confirms the first hypothesis. It will be noted that this interaction effect is in the opposite direction to that to be expected on the basis of a simple regression effect: the resistance conferred by the supportive defense is not only less to immediate attacks, but such as it is, it also decays more rapidly than the greater immediate resistance conferred by the refutation.

There is an alternative theoretical interpretation of this superior persistence of the immunity conferred by the refutational defense, an explanation related to the "sleeper effect" described by Hovland and his colleagues (Hovland, Lumsdaine, & Sheffield, 1949, Ch. 7; Hovland & Weiss, 1951; Kelman & Hovland, 1953). According to these



theorists, if a persuasive message is initially accompanied by a discounting cue, its opinion change impact might actually increase with time passage, or at least decline relatively slowly (Weiss, 1953) as compared with a message not so accompanied. The recall of the persuasive content does, of course, decay over time but so does the recall of the discounting cue, thus, reducing or even reversing the net decay of induced opinion change. In the present situation the refutational defense could be interpreted as containing a discounting cue, namely, mention of the refuted counterarguments, the forgetting of which dampens the decay of the initially induced opinion change. The supportive defense contains no such incidental discounting cue, so that its greatest impact should be felt immediately and decay thereafter without any mitigating effect of a simultaneously decaying discounting factor. Some credence is given to this interpretation by the results of an earlier study (McGuire & Papageorgis, 1961) indicating that the direct strengthening effect prior to any attack was somewhat greater

( $.05 < p < .10$ ) with the supportive than the refutational defense, even though the latter conferred more resistance ( $p < .01$ ) to an attack 2 days later.

*Relative resistance after refutation of same and of alternate counterarguments.* As can be seen in Figure 1b, the resistance conferred by the refutational-same defense, i.e., the defense involving prior refutation of the very counterarguments to be used in the attack, declines monotonically as the interval between defense and attack increases. The decline from the immediate to the 2-day interval is not significant but that from 2 days to 1 week is significant on the .01 level. A quite different, nonmonotonic time trend can be seen with the refutational-different defense, i.e., the defense involving prior refutation of counterarguments different from the ones that are actually to be used in the attack on the given belief. Although this type of defense was inferior ( $p < .05$ ) to the refutation-same defense in conferring resistance to the immediate attack, it has become trivially superior when the attacks do not come until 2

TABLE 2

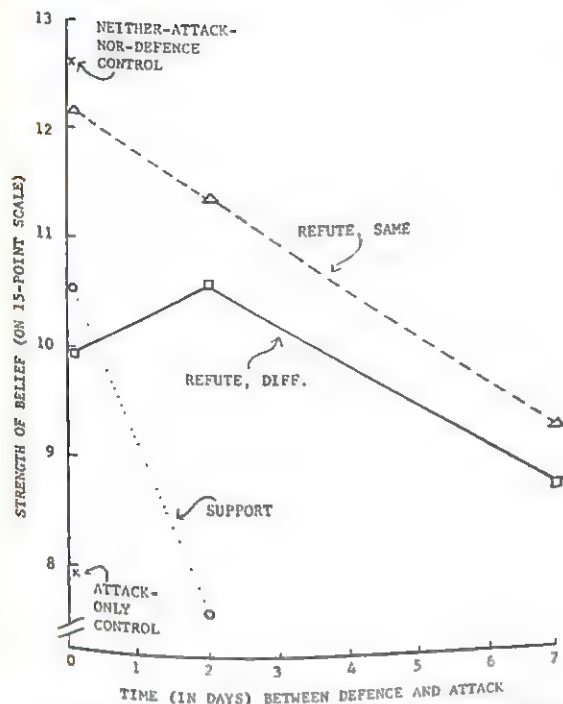
MEAN BELIEF SCORES (on a 15-point scale) AND ANALYSIS OF VARIANCE IN THE CONDITIONS INVOLVING ATTACKS BY THE SAME OR BY NOVEL COUNTERARGUMENTS IMMEDIATELY AFTER AND TWO DAYS AFTER THE REFUTATION DEFENSES, WITH ISSUE-BY-ISSUE SUBMEANS

Issues	Refutational-same defense		Refutational-different defense		All treatments
	Immediate attack	Attack after 2 days	Immediate attack	Attack after 2 days	
Chest X ray	10.10	8.50	8.96	9.92	9.42
Penicillin†	12.59	12.95	11.92	12.48	12.41
Toothbrushing	10.35	9.92	9.84	10.78	10.18
Annual physical	12.41	12.92	10.92	12.62	12.04
All issues	11.36	11.08	10.41	11.45	11.01

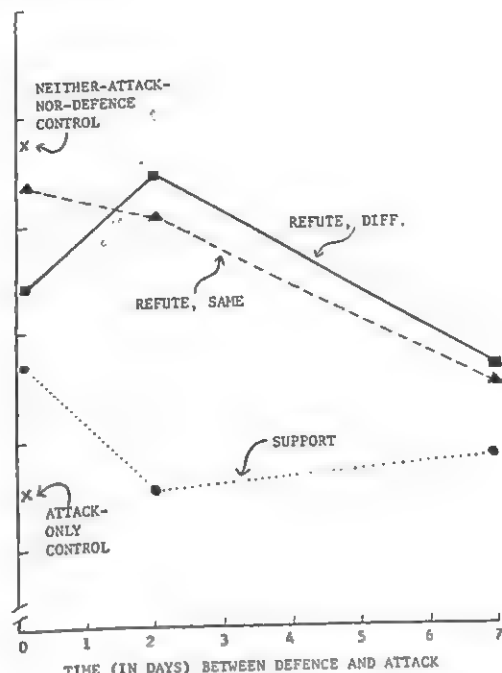
Source	SS	df	MS	F
Type defense (refutational same vs. refutational-different)	248	1	248	2.19
Time (immediate attack vs. 2 days)	119	1	119	1.05
Type × Time	552	1	552	4.88*
Issues	5,965	3	1988	17.59**
Issues × Treatments	517	9	57	0.50
Subject	17,755	79	225	1.99
Residual	16,425	145	113	
Total	41,581	239		

\*  $p < .05$ .\*\*  $p < .001$ .



1a. Based on data from previous experiments. (Zero interval points are based on data from McGuire, 1961a; 2-day interval points, on McGuire and Papageorgis, 1961, and McGuire, 1961b; and 7-day points, on Papageorgis and McGuire, 1961.)

FIG. 1. Persistence of the resistance to persuasion conferred by three types of prior defense: supportive, refutation of the same counterarguments as used in the attack, and refutation of counterarguments different from those used in the attack.



1b. Based on data from the present experiment as shown in Table 1.

days later (see Table 2). The interaction effect between this same vs. different refutational-defense variable and the immediate vs. 2-day interval variable appears on all four issues individually, and the effect combined over issues is significant above the .05 level. Hence, the second hypothesis—that the decay of the resistance conferred against attacks by different counterarguments will be slower than that to attacks by the same counterarguments as had been refuted—is confirmed. As can be seen in Figure 1b, there is actually greater resistance in the “different” refutation condition than in the “same” at both the 2- and 7-day intervals, which is embarrassingly more than the theory demands but this differential is trivial in magnitude.

The results also corroborate the third hypothesis, regarding a delayed-action effect in the resistance conferred by refutation of counterarguments different from those used in the attack. As can be seen in Table 2, the resist-

ance conferred by this type of defense is greater against an attack 2 days later than against an immediate attack on all four issues individually, as well as in the combined results ( $p < .05$ ).

In general, the results from this experiment agree closely with the cross-experimental comparisons from the previous studies in the introductory section. As can be seen from a comparison of Figures 1a and 1b, the two sets of curves are quite similar in shape and even as regards absolute parameters, except that in the present experiment, the refutational-different defense tends to be somewhat more effective than in the previous experiments. The general implication of the study, particularly when considered in the context of the previous studies in this series, is to corroborate further the initial postulate: that the supportive defense confers resistance to persuasion only in so far as the material presented is assimilated and retained—an activ-

ity that the subject is little motivated to carry out in the case of a "truism." The resistance conferred by the refutational defense, on the other hand, derives not only from the assimilation and retention of the bolstering material actually presented but also from the motivational effect of the pre-exposure to threatening material, the mention of the counterarguments, contained in the refutational defense. If this interpretation is correct, then the temporal differentials found among the defenses in this study should be reduced as we move from the truisms used in this study, with respect to which the defense stimulating threat is particularly necessary and possible, to beliefs on more controversial issues. The results do seem fairly general as regards truisms, as can be seen in the trivial magnitude of the Issues  $\times$  Treatments interaction effects (see Table 2).

#### SUMMARY

Theoretical considerations like those which led to the predictions tested in the previous studies of this series on immunizing beliefs against persuasion yielded several hypotheses regarding differential persistence of the immunity conferred by various types of prior belief-defenses. First, it was predicted that the immunity conferred by refutational defenses would decay less rapidly than that conferred by the supportive defenses. Secondly, within the refutational-defense conditions, it was predicted that the conferred resistance to attacks by counterarguments other than the explicitly refuted ones would decay less rapidly than resistance to attacks by the very counterarguments refuted. A related third prediction was that there would be a delayed action effect in the immunity to attacks by novel counterarguments conferred by the refutational defense.

Each of the 160 college students subjects served in two experimental sessions. The first involved reading defensive articles on medical truisms. The defenses involved either arguments supporting the truism, or refutations of counterarguments against the truism, either the very counterarguments to be used in the later attack or alternative counterarguments. The second session came either 2 days (for 80 subjects) or 7 days (for the other 80) after the first, and involved a second defensive treatment on another truism and then attacks on the previously defended and undefended truisms. The subjects' beliefs on all the truisms were then measured. All three hypotheses received substantial confirmation.

#### REFERENCES

- HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. *Experiments on mass communication*. Princeton: Princeton Univer. Press, 1949.
- HOVLAND, C. I., & WEISS, W. The influence of source credibility on communication effectiveness. *Publ. opin. Quart.*, 1951, 15, 635-650.
- KELMAN, H., & HOVLAND, C. I. "Reinstatement" of the communicator in delayed measurement of opinion change. *J. abnorm. soc. Psychol.*, 1953, 48, 327-335.
- MCGUIRE, W. J. The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry*, 1961, 24, 184-197. (a)
- MCGUIRE, W. J. Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *J. abnorm. soc. Psychol.*, 1961, 63, 326-332. (b)
- MCGUIRE, W. J., & PAPAGEORGIS, D. The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *J. abnorm. soc. Psychol.*, 1961, 62, 327-337.
- PAPAGEORGIS, D., & MCGUIRE, W. J. The generality of immunity to persuasion produced by pre-exposure to weakened counterarguments. *J. abnorm. soc. Psychol.*, 1961, 62, 475-481.
- WEISS, W. A "sleeper" effect in opinion change. *J. abnorm. soc. Psychol.*, 1953, 48, 173-180.

(Received March 21, 1961)



# SCHIZOPHRENIC AND NORMAL RESPONSE PATTERNS TO "AVERSIVE" AND "NEUTRAL" ASSOCIATIONS IN TWO PAIRED-ASSOCIATE PARADIGMS<sup>1</sup>

MICHAEL GLADIS

*Veterans Administration Hospital, Pittsburgh, Pennsylvania*

AND GEORGE J. WISCHNER

*University of Pittsburgh*

This study had two principal objectives: to compare the acquisition of stimulus-response associations involving neutral and aversive response members; to study the acquisition of such associations as a function of two methods of paired-associate learning. The two paired-associate paradigms included the conventional paired-associate task (PAC) in which the subject is presented with a stimulus word and has to anticipate the correct response term, and a modified paired-associate paradigm (PAD) described by Ramond (1953) which necessitates a discrimination by the subject. The PAD technique is described in greater detail later.

Two recent studies (Jacobs, 1955; Laffal, 1952) suggest that it is more difficult for college subjects to establish associations involving disturbing or aversive material than those involving neutral material. Laffal, for example, reported a significant difference at between the .05 and .01 levels between the learning of disturbing and neutral items. The absolute mean difference in trials to learn was relatively small (1.19). It may be noted that in these studies, the so-called emotional or disturbing words were specially selected on the basis of such criteria as reaction time in a prelearning word association task, GSR, or type of association. That is, the critical words were idiosyncratic and differed for each subject. Although in both studies the critical words served as response members in a conventional paired-associate task, the stimulus members of the items differed. Laffal em-

ployed picture stimuli, Jacobs used nonsense syllables. In the present study both stimulus and response words were the usual two syllable adjectives selected from the commonly used calibrated lists including the Haagen (1949) and Melton and Safir (Hilgard, 1951) materials.

The present study extended the range of subjects to include not only college subjects but also Veterans Administration hospitalized schizophrenics and Veterans Administration nonpsychotic patients. It should be pointed out that interest centered not on a comparison of levels of performance as between the college and hospital subjects, but on the intra-group learning patterns relative to the classes of words employed.

The inclusion of hospital patients was suggested by the following considerations. It may be noted that even with the utilization of specially selected materials for each subject, the absolute difference in the learning of emotionally toned and neutral associations by college subjects has been relatively small. In recent studies involving schizophrenics, not necessarily learning studies, the findings have been interpreted as suggesting that schizophrenics are peculiarly sensitive to aversive content (Dunn, 1954) and that they tend to avoid aversive personal associations (White, 1949). This led to the expectation that schizophrenics particularly should have difficulty in acquiring associations involving aversive or affectively toned words. The decision to study the learning method variable was stimulated by the rather unusual findings (described later) obtained with PAD, the learning method initially employed.

<sup>1</sup> A portion of this paper was presented at the 1959 Annual Meetings of the Eastern Psychological Association.

TABLE 1

LEARNING MATERIAL WITH CORRESPONDING AVERSIVE-NEUTRAL AND NEUTRAL-NEUTRAL PAIRS

Neutral-Aversive pairs (A-N <sub>A</sub> )		Neutral-Neutral pairs (N-N)	
Stimuli	Response	Stimuli	Response
EVEN LEVEL	FILTHY KINGLY	EQUAL ALIKE	HOPEFUL COVERT
UPPER HIGHER	GUILTY DISTINCT	PLEASANT CHEERFUL	HUMBLE SINCERE
THOROUGH COMPLETE	EVIL SPECIAL	CONSTANT STEADY	ALIVE FAITHFUL
READY PREPARED	HATEFUL NOVEL	TOTAL ENTIRE	FROSTY PROFOUND

## METHOD

*Subjects*

One hundred and twenty subjects<sup>2</sup> divided into six groups of 20 participated in the study. Three groups (60 subjects) were run with PAD and corresponding groups were run with PAC. The groups employed with PAD had the following characteristics:

1. Schizophrenic subjects were male patients in a Veterans Administration hospital diagnosed as the paranoid type in partial remission. Their chronological age range was from 30 to 46 with a mean of 37.7 years.

2. A Veterans Administration hospitalized nonpsychotic group was selected from a Veterans Administration general medical and surgical hospital. This group was convalescing from a fairly wide range of physical ailments, for example, hernia surgery, pneumonia, hypertensive vascular disease. Mean age was 36.3 with a range from 19 to 50 years.

3. College subjects were male students enrolled in an evening graduate psychology course. They ranged from 25 to 48 years with a mean age of 31.8.

The subjects run with the PAC technique were drawn from the same sources as the above subjects and had the following characteristics: The schizophrenic group age range was from 24 to 45 years with a mean of 35.8 years; the nonpsychotics ranged in age from 23 to 50 with a mean of 36.5 years; the college subjects ranged in age from 21 to 52 with a mean of 32.5 years.

The hospital groups on the basis of their WAIS vocabulary scores ranged in intelligence from dull normal to superior. The mean intelligence of these groups would place them in the low average to average range. No comparable data were available for the college subjects.

<sup>2</sup> Appreciation is expressed to the Managers and Staff of the Veterans Administration Hospitals, Leech Farm Road, and University Drive, Pittsburgh, for making space and subjects available.

*Learning Material*

**PAD.** The material consisted of eight item pairs constituting a list totaling 16 items. A stimulus word and two response alternatives composed an item. The subjects' task was to select the correct alternative. The stimulus words for each item pair were synonyms while the response alternatives were identical. For each stimulus a different response alternative was correct. In the following example of an item pair the arrow points to the correct response:



In four of the item pairs the response alternatives included an aversive and neutral word (A-N<sub>A</sub> pairing). In the other four pairs both response alternatives were considered neutral (N-N pairing). Stimulus words were so selected that the association value between stimuli of a pair of items as calibrated by Melton and Safir (Hilgard, 1951) was equal for an A-N<sub>A</sub> and a corresponding N-N pair. Similarly the response alternatives were equated for frequency according to the Thorndike-Lorge (1944) list. Corresponding A-N<sub>A</sub> and N-N item pairs are listed opposite each other in Table 1. In Table 1 the stimulus and response words for the A-N<sub>A</sub> pairs are in the left half and the stimulus and response words for the N-N pairs are in the right half. The four aversive words used were FILTHY, GUILTY, EVIL, and HATEFUL. The judgment of aversiveness was made on an a priori basis by the experimenters and was checked informally with other staff psychologists.

**PAC.** The same learning list was used in the PAC paradigm. An item now consisted of a stimulus word and one response word which had to be anticipated. Each stimulus word was paired with the same correct response word as in PAD. The 16 stimulus response pairs in Table 1 now again constitute these individual items.

*Apparatus and Procedure*

For both phases of the study the learning material was presented on a Hull-type memory drum. There were five different orders of the list. For PAD each item presentation consisted of the simultaneous appearance in the aperture of the drum of both the stimulus and two response alternatives. Each item was exposed for 4 seconds with an intertrial interval of 8 seconds. In the PAC phase the stimulus word appeared in the aperture for 4 seconds, then the stimulus and response words appeared together for 4 seconds. The intertrial interval again was 8 seconds.

In the PAD paradigm this general procedure was employed. After the subject was seated before the apparatus a set of instructions was read aloud to him. In general the subject was informed of the nature of the experiment and then was told that he would see a succession of items of one stimulus word on the left and two response words on the right and

that his task was to learn in each instance which of the two words on the right was correct. He was to indicate his choice by calling out the response word he thought was correct. If the subject called out the correct choice the experimenter said "right" and if the subject's choice was incorrect the experimenter remained silent. The subject was encouraged to guess if he was not sure. The experimenter was seated behind the table on which the memory drum was mounted in full view of the subject. It was felt that an experimental situation in which the experimenter was in view would be less anxiety provoking for the schizophrenic group than one in which he was hidden behind a screen. A learning session included the following: the subject was first given eight trials on a four item practice list of two syllable nouns, 1 minute rest period, 15 trials on the test list, a 2 minute rest period, 15 more trials on the test list. Each subject then received 30 trials unless he reached a criterion of three perfect repetitions prior to 30 trials. For PAC the subjects were given the usual instructions for paired-associate learning. Otherwise the procedure was the same as with PAD.

### RESULTS

Performance curves for the three groups of subjects on the different word classes for PAD and PAC are shown in Figures 1 and 2, respectively. Percentage of correct responses is plotted against five trial blocks.

Figure 1 indicates that the pattern of performance on the various word classes for the two hospital groups is remarkably similar and differs quite sharply from that of the college subjects. The top curves for the hospital groups depicting the courses of learning for the  $N_A$  words (neutral paired with aversive) begin high at approximately 70% correct responses but remain essentially flat showing little change over trial blocks. The middle curves for the hospital subjects are for the two subgroups of correct neutral words combined (neutral paired with neutral). These begin at approximately 50% correct as would be expected on the basis of initial chance performance but also show little evidence of improvement in performance. The bottom curves are for the A words and show a very low rate of initial correct responding, performance approximating 20% correct. This is clearly below chance and certainly far below the corresponding points for the upper curves. An interesting feature of this performance picture is that despite the initial low level of responding and the overall inferiority of performance on the A words as compared with the others,

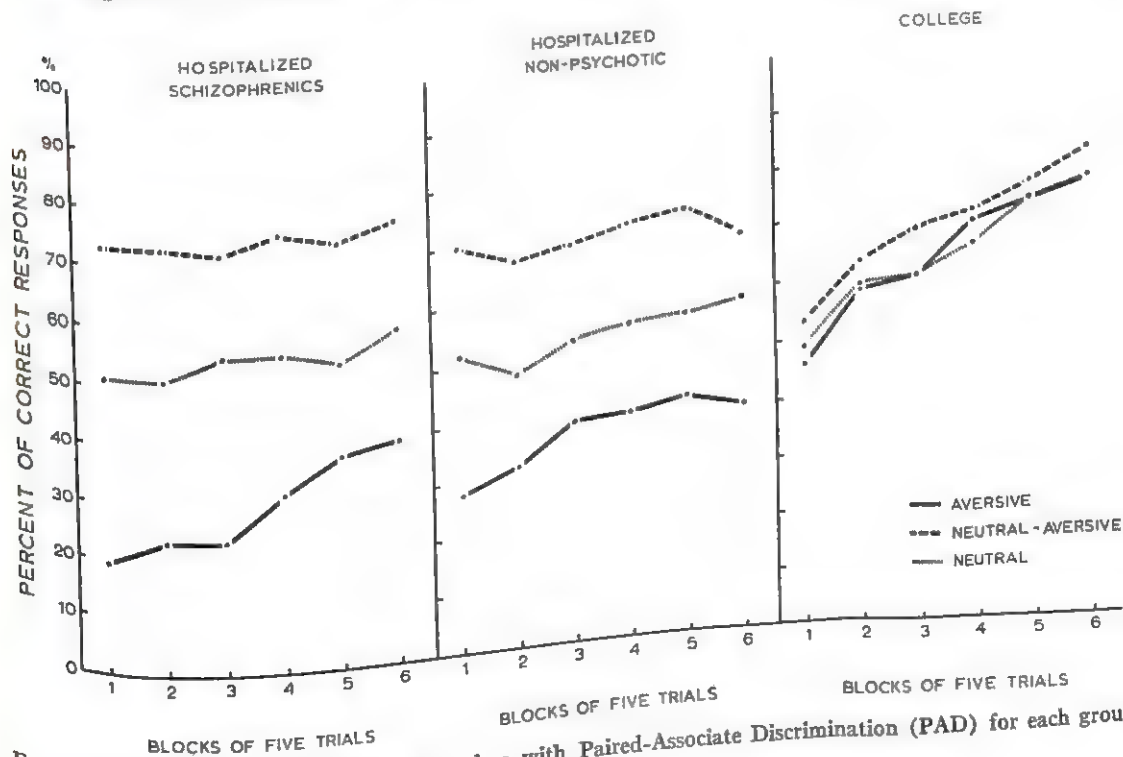


FIG. 1. Course of learning for each word class with Paired-Associate Discrimination (PAD) for each group of subjects.



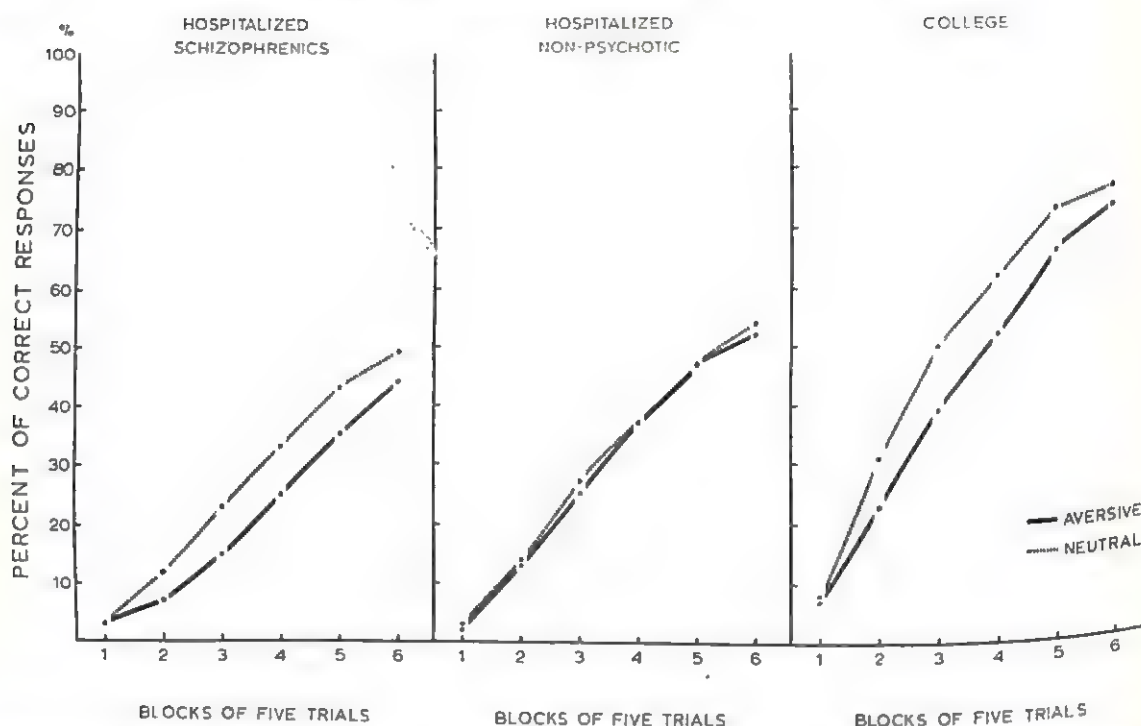


FIG. 2. Course of learning for each word class with Paired-Associate Conventional (PAC) for each group of subjects.

there is some evidence of improvement with a slow but steady rise to about the 40% level. In rather marked contrast the curves for the college subjects do not show the dramatic distinctness evidenced in those for the hos-

pital groups. They all begin at approximately 50% and climb to roughly 80% for the last block of trials. There is clear evidence of learning by college subjects on all word sets. It may be noted that the curve for the A words does tend to fall somewhat below that of the N<sub>A</sub> words, but that it also is not clearly discriminable from the N curve.

TABLE 2

MEAN NUMBER OF CORRECT RESPONSES AND THEIR STANDARD ERRORS MADE BY EACH GROUP ON THE DIFFERENT WORD SETS FOR EACH LEARNING TECHNIQUE

	Schizophrenics		Nonpsychotics		College	
	M	SE	M	SE	M	SE
Paired-Associate Discrimination (PAD)						
Aversive	33.80	5.23	44.05	4.91	75.85	3.22
Neutral	88.50	5.05	84.10	4.00	82.25	3.50
Aversive-Neutral	64.41	3.56	64.27	2.33	76.12	3.42
All Words	250.60	6.37	256.70	8.03	310.35	12.65
Paired-Associate Conventional (PAC)						
Aversive	25.85	4.85	35.15	4.42	51.65	5.87
Neutral	32.68	4.87	36.19	3.08	60.02	4.57
All Words	123.90	18.61	143.80	12.29	231.85	16.49

Figure 2 which shows corresponding data for PAC indicates a greater similarity of intralearning pattern among the three groups of subjects. Here it will be recalled there are only two curves, for A and N words, since there could be no neutral words paired with aversive ones in this paradigm. All groups now showed clear evidence of learning on all word classes. Interestingly both the schizophrenic and college groups tended to do worse on A than on N words, a difference not suggested for the hospital nonpsychotics.

Table 2 presents the mean number of correct responses and their SE made by each group on each word class for the two learning paradigms. The data for PAD include the means for three word classes based upon the four aversive words (A), the four neutral

words paired with them ( $N_A$ ), and the eight neutral words paired with each other ( $N$ ). The PAC data include only means for two word classes, one based upon the four aversive words and the other upon 12 associations involving neutral response words. The mean values in Table 2 for the  $N$  words for the PAD data represent the mean of one-half the number of  $N$  words each subject got correct while the values for the  $N$  words in the PAC data represent the mean of one-third the number of  $N$  words each subject got correct. The bottom row under each technique presents the mean number of correct responses and the  $SE$  for all words combined, that is, for the entire list of 16 items.

Separate analyses of variance were performed on the data for each learning technique. The results are summarized in Table 3. The analysis of variance for PAD indicates highly significant differences between groups and between word sets. The Groups  $\times$  Word Set interaction was also highly significant. It will become clear that these findings are attributable largely to the differences between the college subjects on the one hand and the two hospital groups who were quite similar in performance, on the other.

On the entire list, utilizing the PAD procedure, it is evident from Table 2 that the college subjects achieved a considerably higher number of mean correct responses (310.35) than either the schizophrenics (250.60) or the nonpsychotics (256.70). These differences between the college subjects on the one hand and the hospital groups on the other were significant at the .01 level. The general superiority of the college subjects is not too

surprising and as stressed earlier, it was not a primary objective to compare the college with other subjects. More noteworthy perhaps is the similarity in performance of the schizophrenic and nonpsychotic groups.

Considering now PAD performance on the word classes it is evident that the schizophrenics performed far less well on the  $A$  words than on the  $N_A$  words. The very large mean difference of 54.7 between these word classes was highly significant. The nonpsychotic hospital subjects showed a remarkably similar pattern with a much lower number of mean correct responses on the  $A$  words (44.05) than on the  $N_A$  words (84.10). College subjects also tended to do less well on the  $A$  than on the  $N_A$  words but the difference between 75.85 for the former and 82.25 for the latter, is far less dramatic. In this connection it must be noted that for these subjects performance on  $A$  words was highly similar to that on the  $N$  words. In contrast both hospital groups performed significantly better on the  $N$  than on the  $A$  words.

Turning now to the PAC data, Table 3 indicates that the between groups  $F$  was significant at the .01 level; the  $F$  between word classes now was significant at only the .05 level. The interaction here was not significant in contrast to that for the PAD data. Table 2 indicates again that college subjects generally performed at a higher level than the hospital groups which again were more similar to each other. Although the latter groups attained fewer mean correct responses on the  $A$  than on the  $N$  words the difference for either group was far smaller than it was for PAD. In fact the college group showed a greater mean dif-

TABLE 3  
SUMMARY OF THE ANALYSES OF VARIANCE FOR THE PAIRED-ASSOCIATE DISCRIMINATION (PAD)  
AND PAIRED-ASSOCIATE CONVENTIONAL (PAC) LEARNING SCORES

Source	PAD			PAC		
	df	MS	F	df	MS	F
Between groups	2	4,547.12	14.99**	2	7,705.45	10.96**
Between subjects	57	303.34		57	703.15	
Between word sets	2	16,885.30	56.92**	1	885.63	6.47*
Word $\times$ Groups	4	3,082.80	10.39**	2	150.43	1.10
Pooled subjects $\times$ Words	114	296.64		57	136.80	
Total	179			119		

\*  $p < .05$ .  
\*\*  $p < .01$ .

ference between these two word classes (51.65 vs. 60.02) than did the schizophrenics (25.85 vs. 32.68). The nonpsychotic subjects showed no difference in performance on these two word sets (35.15 vs. 36.19). For the schizophrenics, however, the difference just approached significance at the .05 level; for the college subjects the difference was not significant at this level. Although the schizophrenics achieved a lower mean number of correct responses on the A words (25.85) than did the nonpsychotics (35.15), this difference was not statistically significant.

Certainly performance difference between word classes was much greater with PAD than with the PAC technique.

#### DISCUSSION

The overall findings of the present study tend to support a conclusion that associations involving aversive response terms tend to be learned less well than associations containing relatively neutral terms. However, more specific considerations indicate the need for qualification of such a conclusion and point to the operation of at least two important variables that may affect such learning. These include the learning method and Veterans Administration hospital status.

Among the more dramatic findings of the present study were the unique intralearning patterns on the various word classes displayed by the hospitalized groups with PAD and the marked differences in performance patterns obtained with the two learning paradigms. The PAD data for both hospital groups tend to support the conclusion that not only schizophrenics but Veterans Administration patients generally may be particularly sensitive to material which they react to as aversive. Why the aversive words of this study, selected as they were from conventional learning materials, should be as aversive as they seemed, is not readily clear. It is worth emphasizing that if only schizophrenic subjects had been employed one might have been willing to conclude that the results obtained support a hypothesis previously mentioned, namely, that schizophrenics are uniquely sensitive to aversive stimulus characteristics in the environment. The markedly similar performance of

the nonpsychotic subjects strongly suggests that Veterans Administration hospital status and not diagnostic label was the relevant factor in this study.

Moreover, why the apparent sensitization to A words should have shown up so differently in the two learning situations raises the question of what differences in the paired-associate techniques may have been related to such results. With PAD patients tended from the very beginning to show evidence of responsiveness to the aversive alternative in the aversive-neutral pairs. This is not to say that they never verbalized the aversive words. Their performance of only 70% correct on the N words suggests the contrary. This 70%, it is to be noted does not represent improvement in performance as it will be recalled that patients showed no apparent learning on the  $N_A$  items. To interpret this absence of learning as being due partly or solely to the fact that these neutral words were paired with aversive ones seems to be contraverted by the absence of learning on the neutral-neutral pairs. A more important related finding is that despite the performance, much below chance, by patients on the A words, differential reinforcement was beginning to be effective as indicated by the evidence of significant improvement in performance only on these words. The importance of the learning method variable is further emphasized by the contrasting result for PAC which showed that hospital subjects were clearly capable of a significant amount of learning on exactly the same associations employed with PAD.

An analysis of the PAD and PAC paradigms suggests that differences in the visual display that each technique presents to the subject may be related to the different results obtained with the two learning techniques. In the case of PAD the subject had to choose between two response alternatives presented simultaneously. This may have afforded the subjects a greater opportunity to observe differences between the response alternatives and assisted them in ordering the response words to subclasses which evoked differential reactions in them. The PAC procedure which required the subject to pull a single word out of 16 possible ones did not readily allow for such comparison and possible ordering of re-



sponse terms into subclasses. This is not to suggest that the differences in visual display represent the only relevant dimensions on which the two learning methods differ. The findings point up the need for further analysis and study of other factors which differentiate the methods, for example, method of reinforcement, recognition vs. recall, etc. In relation to the latter it is interesting that the hospital subjects at least showed more evidence of learning with PAC which requires recall than with PAD which is based more on recognition. This would appear to be contrary to the usual findings indicating more efficient performance with recognition than with recall.

Reinforcement in PAD was relatively unable to overcome the reaction aroused by the visual display. The general absence of learning even on the N-N word sets may have resulted from some complex interaction associated with the inclusion in a single list of all word classes. This suggests the desirability of a study which utilizes different lists, each composed of only one word class. Another consideration is that PAD required a discrimination and that insofar as intelligence may be related to this kind of performance, patient subjects who were of lower intelligence than college subjects could not learn with PAD as they did with PAC. However, analysis indicated that there was no significant relationship between performance and intelligence measures for either learning method.

So far, attention has been focused on the hospital subjects. It is significant that even the college subjects showed a trend to perform less well on associations involving aversive responses than neutral responses. These differences were much smaller and more in accord with those reported in previous studies. On PAD the difference between aversive and neutral associations was significant but it is well to note that the performance on A and N words was quite similar. This suggests that a mechanism similar to that operating in patients may also have been operating for the college subjects in the case of A-N<sub>A</sub> items. Obviously it was not as clear for the college as for the hospital subjects. Moreover with both learning techniques, college subjects clearly showed improvement in per-

formance attaining approximately the same final level of 80% correct in both cases.

In conclusion, this study points up several important considerations: (a) the particular method of paired-associate learning may be a relevant factor and one cannot assume equivalence for all classes of subjects of PAC and PAD; (b) hospital status may be an important variable in such studies as the present one and it would appear relevant to consider the extent to which findings in the present study are unique to Veterans Administration or generalizable to non-Veterans Administration hospital and other subjects; (c) in so far as aversiveness may be a relevant dimension in learning studies generally, it seems important to scale on this dimension materials taken from frequently used sources.

#### SUMMARY

The objectives of this study were to compare the acquisition of stimulus-response associations involving neutral and aversive response members; to study the acquisition of such associations with a conventional paired-associate paradigm (PAC) and a paired-associate paradigm involving discrimination (PAD). Veterans Administration hospitalized schizophrenics, Veterans Administration hospitalized nonpsychiatric patients, and college subjects learned the same 16 associations with the two different methods of learning. The results in general supported the conclusion that it is more difficult to form associations involving aversive words than neutral words. Both hospitalized groups showed a marked sensitivity to the aversive words with PAD and their pattern of performance was quite different from that of the college subjects. On the other hand the pattern of performance for all three groups with PAC was much more similar and the sensitivity to the aversive words was much less marked with the hospital groups. The results point up the following considerations as important: (a) the particular method of paired-associate learning used can be a relevant factor and one cannot assume equivalence of PAD and PAC with all classes of subjects; (b) hospital status may be an important variable in such studies and it would appear pertinent to consider the extent to which the findings

of the present study are unique to Veterans Administration subjects or generalizable to non-Veterans Administration hospital and other subjects; (c) since the aversive words used in the present study were taken from frequently used sources for learning material, it might be valuable to scale such material on a dimension of aversiveness.

## REFERENCES

- DUNN, W. L., JR. Visual discrimination of schizophrenic subjects as a function of stimulus meaning. *J. Pers.*, 1954, 23, 48-64.
- HAAGEN, C. H. Synonymity, vividness, familiarity and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, 27, 453-463.
- HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517-567.
- JACOBS, A. Formation of new associations to words selected on the basis of reaction-time-GSR combination. *J. abnorm. soc. Psychol.*, 1955, 51, 371-377.
- LAFFAL, J. The learning and retention of words with association disturbances. *J. abnorm. soc. Psychol.*, 1952, 47, 454-462.
- RAMOND, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, 46, 120-124.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, Bureau of Publications, 1944.
- WHITE, MARY ALICE. A study of schizophrenic language. *J. abnorm. soc. Psychol.*, 1949, 44, 61-74.

(Received February 14, 1961)

## SOME EFFECTS OF SHARED THREAT AND PREJUDICE IN RACIALLY MIXED GROUPS<sup>1</sup>

EUGENE BURNSTEIN

*Michigan State University*

AND

ADIE V. McRAE

*University of Texas*

When members of a social system are threatened, marked changes seem to occur in social relationships (Jacobson & Schachter, 1954; Schachter, Nuttin, de Monchaux, Maucorps, Osmer, Duijker, Rommetveit, & Israel, 1954). Where the consequences of the threat and the responsibilities for coping with it are shared, an increase in group cohesion and a reduction in disruptive antagonisms may occur (French, 1941; Leighton, 1945; Pepitone & Kleiner, 1957; Sherif & Sherif, 1953; Wright, 1943).<sup>2</sup> The application of this general finding to the study of particular social problems can have important consequences. If the social system in question is a society, community, or group containing distinct religious or racial subgroups, concern about a shared threat may lead to a decrease in the amount of hostility expressed toward these minorities.

In the first explicit attempt to test the hypothesis that shared threat reduced social prejudice, Feshbach and Singer (1957) presented a set of questions to individuals de-

signed to provoke concern about dangers which confront the community as a whole, e.g., floods, hurricanes, atomic attack. Immediately afterward a social prejudice questionnaire was administered. Responses on the final questionnaire were compared to those the person made a month earlier. The authors reasoned as follows:

Under the impact of a common threat . . . one's reference group may become the population that is subjected to the danger. If this reference group now includes both Negro and white, whereas under ordinary stimulus conditions the reference group has been primarily the white population, then the social distance between white and Negro should decrease, with a corresponding decrement in social prejudice (p. 412).

The results gave only weak support to the hypothesis. However, there are considerations which suggest the shared threat induced by this method may have been relatively weak.

Requiring people to think about a community-wide disaster does not insure that they view it as one in which the suffering and responsibilities are equally distributed among all community members. In a pilot study conducted by the senior author, 47 male students in the elementary psychology course at the University of Texas were administered the first four of the five "Flood and Hurricane Threat questions" from Feshbach and Singer (1957). In addition they were asked if such a disaster struck Austin, Texas, would all or nearly all socioeconomic levels, ethnic groups, or neighborhoods be equally affected. Only 27% thought this to be likely. Over 30% thought that there would be large differences among various groups in the degree to which they suffered from such disasters. Similar differences occurred in regard to the distribution of the burden for coping with the disaster. Therefore, given this method of induction, the extent to which the subjects perceived the threat to be shared is ambiguous.

Furthermore, in a highly complex social

<sup>1</sup> This research was supported in part by the United States Air Force under Contract No. AF 49(638)-460 monitored by the Air Force Office of Scientific Research of the Air Research and Development Command and the University of Texas Research Institute.

<sup>2</sup> A shared threat has also been observed to increase hostility among group members. In Nazi concentration camps, inmates went so far as to identify themselves with the source of the threat (Bettelheim, 1943; Cohen, 1953). At present it is not completely clear what are the necessary and sufficient conditions for a shared threat to reduce intermember hostility. However, a review of the literature suggests the important determinants are (a) the overwhelming nature of the threat, (b) the degree to which group action can ameliorate the threat, and (c) the degree to which members equally share the consequences of the threat and the responsibilities for coping with it. In the concentration camp the threat was quite overwhelming. Group action provided little amelioration; in fact, for many inmates a reduction in threat was only possible by dissociating themselves from the group. Treatment varied with the category of the inmate, and little role differentiation occurred other than imposed by the camp administration.



system such as a community, multiple group membership is the rule. During a disaster, the person may experience severe role conflicts. In spite of the perception that the threat is shared equally by all community members, the role of a father, neighbor, or plant manager may be more salient than that of community citizen. This phenomenon is vividly documented by Killian (1952) in his study of the Texas City explosion and of three tornado-torn towns in Oklahoma. Thus, even when a shared threat is perceived to exist in a community setting, it is uncertain whether the community as a whole or some subsystem will become the salient reference group. In the latter case, minorities within the community remain outgroups in terms of the social relations which are salient for the person at that time. Under such conditions, social prejudice may be unaffected.

In order to test the hypotheses that shared threat reduces the expression of hostility toward minorities either one of two general procedures can be used to minimize these processes which vitiate the threat induction: some method may be introduced to assure that the person faced by a community-wide threat takes the community as the salient reference group, or the threat may be induced in a simpler social system in which the number of group memberships available to the person is sharply reduced. Both procedures attempt to decrease the likelihood that roles or reference groups external to the threatened social system become salient. The present experiment utilizes the second method. Members of racially mixed groups cooperate to solve a logical problem. In these groups, failure is clearly shared by all members. At the same time all members have a role in coping with the status loss that results from failure (Deutsch, 1953). The social system, furthermore, is simple enough so that under the threat of status loss few, if any, alternative roles are likely to become salient other than membership in the particular problem solving group.

Another source of variation in the expression of hostility toward an individual Negro that should be controlled is the attitude of the other members toward this racial group as a whole. The stronger the person's anti-Negro attitudes, the more likely is he to be

hostile toward a Negro member of his problem solving group. Thus, in the present study anti-Negro attitudes as well as shared threat will be examined.

If the expression of hostility toward a Negro group member varies directly with the strength of anti-Negro attitudes and inversely with the degree of shared threat, then the following predictions can be made: (a) high prejudiced individuals under nonthreatening conditions will express the greatest amount of hostility toward the Negro member; (b) low prejudiced individuals under shared threat will express the least amount of hostility toward the Negro member; (c) high prejudiced individuals under shared threat and low prejudiced individuals under nonthreatening conditions will display an intermediate amount of hostility toward the minority group member. In the situation under study hostility may be manifested in direct evaluations made of the Negro, in the frequency with which the Negro is rejected from the group, and in the avoidance of communication with him during the problem solving interaction.

## METHOD

*Subjects and confederate.* Forty-eight male students in the elementary psychology course at the University of Texas were used as subjects. Participation fulfilled a course requirement. Several weeks before the experiment they were assessed as to their level of anti-Negro prejudice by means of Holtzman's D scale (Kelly, Ferson, & Holtzman, 1958), in the form of a "Student Attitude and Opinion Questionnaire." This was administered by the instructors in a number of the sections of the course. The distribution of prejudice scores was split at the median; subjects falling above the median were considered high in prejudice, those below the median, low in prejudice. In order to minimize the possibility of prior acquaintanceship, the four subjects used in each experimental group were drawn from separate sections.

A Negro confederate was paid to serve as a member in all experimental groups. The four other members were, in one half of the groups, all high prejudiced subjects, in the other half, all low prejudiced subjects. The confederate participated in several pilot groups to attain maximum familiarity and skill with the type of problem to be used in the experiment. It was necessary to tell him about all phases of the experiment and its objectives. The only information that was withheld from him was the extent of prejudice of the subjects with whom he was to work.

*Procedure.* Six groups were run with low prejudiced subjects and six with high prejudiced subjects. Within each of these two conditions of prejudice,

shared threat was induced in half of the group, while a nonthreatening or successful state of affairs was induced in the other half. The design, therefore, consisted of three groups of four subjects, plus the confederate, under each of the following conditions: High Prejudice, Nonthreat (HPNT); Low Prejudice, Nonthreat (LPNT); High Prejudice, Threat (HPT); and Low Prejudice, Threat (LPT).

Communication among the subjects occurred around a table similar to that used by Leavitt (1951). The subjects were seated so that each was separated from the next by a vertical partition extending from a post in the center of the table. The center post had slots allowing subjects to push written messages to other members. Direct communication was permitted among all members. Messages were written on colored cards corresponding to the color of the cubicle from which each subject operated.

As each subject arrived he was given a seat in front of his cubicle. When all subjects had taken their places, they were asked to stand and see who the other members were but not to engage in any conversation. A copy of the instructions was given to each member and they were asked to follow as the experimenter read them aloud. In summary form, the instructions were as follows:

The purpose of this procedure is to evaluate how groups work together in solving problems when communication is limited to written messages. It has been found that a procedure such as this can be used to single out groups with different levels of skillfulness, efficiency, and creativity. The university recently has become quite interested in estimating how productively undergraduates can work together in groups. They have suggested that the Psychology Department initiate this program of evaluating groups of students with respect to these qualities. Thus, a record will be kept for the university administration of the performance of the group participating in this preliminary testing. Skillful, efficient, and creative group problem solving will be reflected in the time that it takes the group to solve the problem, i.e., how long after starting before every member has the correct answer. Each member will receive a grade that is based on how well his group performs in solving these problems. This means, of course, that everybody in the group gets the same grade. The grade a group receives will depend on how its performance compares to that of a large number of other groups of college students in Texas who have worked on the same type of problem in the same type of situation.

During the reading of the instructions the subjects were standing facing each other.

All groups were given four successive problems to solve—Leavitt's (1951) "common symbol" problem. They were instructed that each member had been given a different set of symbols and that their task, as a group, was to discover the symbol that was common to all members. When a member knew what this symbol was, he was to put it on a white slip

and place it on top of his section of the center post. The group was considered to have completed the problem when all members had placed their white slip on the center post.

At the conclusion of Task 2, subjects were told to stand, stretch their legs, but not to converse. They were seated and given an evaluation of their performance. Half of the high prejudiced groups and half of the low prejudiced groups (HPNT and LPNT) were told that their performance was well above average. The remaining high prejudiced and low prejudiced groups (HPT and LPT) were informed that they had performed poorly compared to the average performance of similar groups. The experimenter reinforced these evaluations by making two or three positive or negative statements about the group's performance during or immediately after both Tasks 3 and 4. At the end of Task 4, the experimenter similarly evaluated the groups with respect to their overall performance. While the final evaluation was made subjects were standing in front of their cubicle facing each other.

Immediately following the final evaluation a post-questionnaire was administered. On six-point scales, subjects rated the experimenter in terms of his "competence as a psychologist" and in terms of their "liking" for him. Similarly, the test situation was rated for its "fairness," its "worthwhileness" and its "interest." To partially assess the success of the threat induction, subjects were asked to rate how "depressed" they felt at the results of the test. Three items allowed subjects to evaluate the other four members. Two of these items involved ranking members in terms of their contribution to the solutions and in terms of who the subjects liked best. The third item required subjects to rate other members for their estimated "communication and problem solving skill in everyday life." At the end of the questionnaire subjects were given a sheet which asked if they wished to replace one of the present members with a new one from the subject pool at the next testing session. If they did desire to do so, they were to indicate the rejected member by encircling one of the four listed colors which corresponded to the color of his cubicle.

After completing the questionnaires, subjects were given a full explanation of the nature of the experiments.

## RESULTS

An analysis of variance of the mean times required for task completion by the four experimental groups indicates that completion time decreases significantly over trials ( $F = 8.06, p < .001$ ). This is in accord with Leavitt's (1951) findings regarding improvement in performances over successive tasks.

To assess the success of the threat induction, two  $t$  tests were made, one on the self-ratings of depression as a result of the test,



TABLE 1  
EVALUATION OF NEGRO CONFEDERATE

	Mean task contribution rank		Mean likability rank		Mean skill ratio	
	Shared threat		Shared threat		Shared threat	
	Present	Absent	Present	Absent	Present	Absent
Prejudice						
High	2.58	4.33	2.60	3.25	1.02	1.37
Low	2.42	3.25	2.58	2.83	.91	1.17
Difference tested	<i>t</i>		<i>t</i>		<i>t</i>	
HPNT vs. HPT	3.12****		2.57***		2.50***	
HPNT vs. LPT	3.42****		2.16**		2.52***	
HPNT vs. LPNT	1.85*		1.02*		—	
LPT vs. LPNT	1.28*		—		1.33*	
LPT vs. HPT	—		—		—	

\* Not significant.

\* Significant at .10 level.

\*\* Significant at .05 level.

\*\*\* Significant at .02 level.

\*\*\*\* Significant at .005 level.

another on the subjects' ratings of themselves and other white members for their communication and problem solving skills in everyday life. The tests indicated that threatened subjects felt more depressed than nonthreatened subjects ( $t = 2.82$ ,  $p < .01$ ). Similarly, threatened subjects graded themselves and other white members significantly lower in everyday communication skills than nonthreatened subjects ( $t = 3.67$ ,  $p < .001$ ). There were no reliable differences as a function of threat in regard to the subjects' evaluations of the experimenter and of the test situation.

To determine differences in hostile expression resulting from shared threat and prejudiced attitude,  $t$  tests were run on the post-questionnaire items in which the subjects ranked the confederate in terms of contribution to task solutions, liking for him, and in which they estimated his everyday communication and problem solving skill. It was predicted that maximum hostility would be expressed in HPNT conditions; the least in the LPT conditions; while HPT and LPNT subjects would express an intermediate amount. In Table 1 the mean rank for contribution to the solutions given to the Negro confederate are presented. A rank of 1 indicates the great-

est contribution, a rank of 5, the least contribution. The order of these mean ranks correspond exactly to the predicted order. However, only the differences between HPNT and HPT and between HPNT and LPT are statistically reliable. The difference between HPNT and LPNT approaches, but does not reach an acceptable level of significance ( $p < .10$ ). For the mean rank given to the Negro confederate in regard to "liking," a score of 1 indicates the greatest relative liking for the confederate, 4 indicates the least liking. The Negro would be expected to be ranked lowest in the HPNT condition, highest in the LPT condition, and intermediate in the LPNT and HPT conditions. The results show the order of mean ranks once again conform to what was hypothesized. Only the differences between HPNT and HPT and between HPNT and LPT are significant. On the third item, subjects were required to estimate their fellow members' everyday communication and problem solving skill. In the context of this item, hostility may be expressed toward the Negro by rating him lower than the other group members. Ratings of group members, however, were shown to be biased by the presence or absence of threat. This is corrected by using a ratio of the rating given to the Negro by each subject over the mean rating given by the subject to all other members. A high degree of similarity between the Negro's rating and the mean rating is indicated as the ratio approaches 1. A ratio greater than 1 means that the confederate is considered less skillful than the average group member, less than 1 indicates he is considered more skillful than the average. Once again the obtained order fits the prediction exactly. However, only the differences between HPNT and HPT and between HPNT and LPT are significant.

With respect to the more general hypothesis that shared threat reduces the expression of hostility toward the confederate, responses to the above three items were analyzed by  $t$  tests for subjects exposed to shared threat and those not exposed, regardless of prejudice. The mean ranks given to the confederate on the first two items and the mean ratio given on the last item by threatened individuals were 2.50, 2.54, and 0.96, respectively.



The same means for nonthreatened subjects were 3.70, 3.04, and 1.27, respectively. The difference between threatened and nonthreatened subjects on the first item was significant at the .005 level; the difference on the second item was significant at the .05 level; and the difference between the ratios was significant at the .01 level.

If they wished, subjects were given the opportunity to vote privately on rejecting a member from the group. In the HPNT condition 9 of the 12 subjects decided to reject a member. Of these 9 rejections, 6 were of the Negro confederate. With 9 subjects making use of their privilege to reject 1 of the 4 members in their group, it is highly improbable that as large or a larger number of these rejections would be directed toward one member by chance ( $p < .01$ ). In the LPNT condition, 8 subjects wished to reject another member. The confederate received 3 of these rejections. Both in the HPT and LPT conditions 9 subjects decided to reject another member; and in each of these conditions 2 rejections were directed toward the confederate. In none of the latter three conditions did the frequency of rejecting the confederate depart significantly from what would be expected by chance alone. Thus, only under the HPNT condition, where the strongest expression of hostility toward the confederate was expected to occur, is the Negro rejected more frequently than chance.

Another significant source of information concerning the orientation of the members toward the Negro is the proportion of the task messages sent to him during the course of the problem solving interaction. Earlier studies have shown that interpersonal dislike can be coordinated to an increase in the barriers to communication (Festinger, Cartwright, Barber, Fleischl, Gottsdanker, Keysen, & Leavitt, 1948; Festinger, Schachter, & Back, 1950; Potashin, 1946). Thus, it is reasonable to expect that the amount of task communication with the Negro would vary inversely with the degree of hostility felt toward him. The total number of messages each subject sent to all other subjects was counted. The percentage of this total which the subject sent to the confederate was then computed. This was done only for those tasks following the initial in-

duction of shared threat, i.e., Tasks 3 and 4. The analysis of variance of these percentages indicates that only the  $F$  ratio (4.64) for prejudice is significant ( $p < .05$ ). The differences between threat conditions and between tasks did not approach significance. Figure 1 shows that the low prejudiced subjects, both threatened and nonthreatened sent a greater proportion of their messages on Tasks 3 and 4 to the confederate than subjects in either high prejudiced condition.

### DISCUSSION

It appears that the expression of hostility toward a Negro group member varies directly with the strength of anti-Negro attitudes, and inversely with the degree of shared threat. Moreover, prejudice against Negroes as a group may be expressed through a reduction in communication to an individual Negro. This is similar to Schachter's (1960) observations regarding communication to a persistent deviant. Of course, since there could be no question of the Negro changing his "deviant" position, i.e., his status as a Negro, there was no initial rise in communication to the confederate as was found by Schachter during the early phases of interaction. Moreover, it is interesting to note that avoidance of

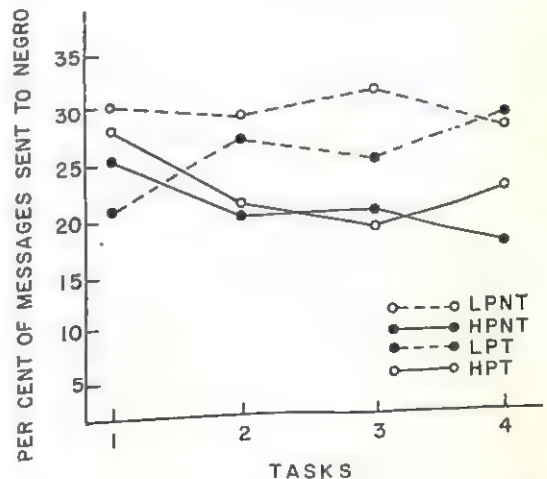


FIG. 1. Percentage of messages sent to Negro confederate by Low Prejudiced, Nonthreatened (LPNT) groups, High Prejudiced, Nonthreatened (HPNT) groups, Low Prejudiced, Threatened (LPT) groups, and High Prejudiced, Threatened (HPT) groups over the four tasks.

communication with the Negro by high prejudiced subjects occurred in a situation where messages were of an impersonal, task oriented nature and where the Negro member possessed information of value to other members in solving the problem.

The prediction that shared threat would inhibit tendencies to avoid communication with the Negro was not confirmed. No difference in communication to the confederate appeared as a function of shared threat. There are a number of possible explanations as to why shared threat reduced the expression of hostility toward the Negro in terms of direct evaluation on the postquestionnaire, but had no effect on the tendency to avoid communication with him. The first bears on the procedure used to induce the threatening and nonthreatening conditions. It will be recalled that the evaluations of the group's performance by the experimenter, which was the means whereby threat was induced, was not made until after the second task. This was relatively late in the problem solving process. The reorganization of the person's initial attitude toward the Negro may take some time. Thus, attitude change may not have occurred in time to appreciably affect task communication. This explanation loses some of its force when one notes in Figure 1 that the differences in communication to the Negro as a function of prejudice occurs more markedly in the second two tasks than on the first two. An ongoing attitude change process should at least prevent such a difference from becoming more pronounced. Nevertheless, it might still be argued that with partitioned cubicles, a relatively long period of time is required before subjects become impressed with the fact that one member is a Negro; and still later, more time is needed for failure and status loss to sink in. Thus, the experiment may have obtained a sample of behavior when the subjects had fully noted the presence of the Negro but before the shared threat had an appreciable effect on communication. On this basis it would be predicted that if more than four tasks were given, subjects in the HPT condition would eventually begin to increase communication with the confederate.

A second line of reasoning assumes that avoidance of communication is a less direct

form of hostile expression than rating the person as poor with respect to certain valued attributes. It also focuses on what is being affected by the induction of shared threat. Are prejudiced attitudes being remolded, or are the expressions of hostility stemming from such attitudes being inhibited without any underlying attitude change? If attitude change had occurred under shared threat, less overall hostility, direct and indirect, should be expressed toward the confederate. This did not occur. If, however, shared threat served to inhibit direct aggression without modifying prejudiced attitudes, our expectation would be quite different. In this case, it would be anticipated that high prejudiced individuals confronted by a common threat would express a smaller amount of direct hostility toward a minority group member than equally prejudiced but unthreatened individuals. Both groups, nevertheless, would be expected to express a similar amount of indirect hostility (avoidance of communication with the Negro) which would be greater than that manifested by less prejudiced individuals.

Finally, discriminatory behavior based on cultural norms and the affective orientation toward Negroes may under certain conditions be uncorrelated. One can follow the discriminatory practices of one's group without necessarily entertaining feelings of hostility. This suggests a third possible interpretation. Since about 75% of the items on the prejudice scale used in this study concern appropriate behavior toward Negroes, a high anti-Negro prejudice score may indicate that the person has strongly internalized the discriminatory behavior patterns of Texas culture. The strength of these norms regarding behavior may not be appreciably modified by a momentary event in a temporary group. Thus, the shared threat induced may have produced a positive change in the affective orientation toward the Negro group member while having no influence on conformity to cultural patterns which stress avoidance of equal status interaction.

#### SUMMARY

The purpose of this experiment was to test the relationship between shared threat and the expression of prejudice hypothesized by Feshbach and Singer (1957). Forty-eight sub-

jects, varying with respect to anti-Negro prejudice, were placed under conditions of shared threat or nonthreat, in task oriented, cooperative work groups. A Negro confederate was a member in each group.

It was found, as hypothesized, that under conditions of shared threat a reduction in the expression of prejudice occurs in terms of direct evaluation of the Negro by other group members on a posttask questionnaire. No significant differences in the amount of communication to the confederate occurred as a result of the threat induction. However, significantly fewer messages were addressed to the Negro by the high prejudiced subjects, regardless of the presence or absence of shared threat.

## REFERENCES

- BETTELHEIM, B. Individual and mass behavior in extreme situations. *J. abnorm. soc. Psychol.*, 1943, 38, 417-452.
- COHEN, E. A. *Human behavior in the concentration camp*. New York: Norton, 1953.
- DEUTSCH, M. The effects of cooperation and competition upon group processes. In D. Cartwright & A. Zander (Eds.), *Group dynamics research and theory*. Evanston, Ill.: Row, Peterson, 1953. Pp. 319-353.
- FESHBACH, S., & SINGER, R. The effects of personal and shared threat upon social prejudice. *J. abnorm. soc. Psychol.*, 1957, 54, 411-416.
- FESTINGER, L., CARTWRIGHT, D., BARBER, KATHLEEN, FLEISCHL, JULIET, GOTTSDANKER, JOSEPHINE, KEYSER, ANNETTE, & LEAVITT, GLORIA. A study of a rumor: Its origin and spread. *Hum. Relat.*, 1948, 1, 464-486.
- FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups: A study of a housing community*. New York: Harper, 1950.
- FRENCH, J. R. P., JR. The disruption and cohesion of groups. *J. Psychol.*, 1941, 36, 361-377.
- JACOBSON, E., & SCHACHTER, S. (Eds.) Cross-national research: A case study. *J. soc. Issues*, 1954, 10(4), 5-68.
- KELLEY, J. G., FERNON, J. E., & HOLTZMAN, W. H. The measurement of attitudes toward the Negro in the South. *J. soc. Psychol.*, 1958, 48, 305-317.
- KILLIAN, L. M. The significance of multiple-group membership in disaster. *Amer. J. Sociol.*, 1952, 57, 309-314.
- LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, 46, 38-50.
- LEIGHTON, A. *The governing of men*. Princeton: Princeton Univer. Press, 1945.
- PEPITONE, A., & KLEINER, R. The effects of threat and frustration on group cohesiveness. *J. abnorm. soc. Psychol.*, 1957, 54, 192-199.
- POTASHIN, R. A sociometric study of children's friendships. *Sociometry*, 1946, 9, 48-70.
- SCHACHTER, S. Deviation, rejection, and communication. In D. Cartwright & A. Zander (Eds.), *Group dynamics research and theory*. Evanston, Ill.: Row, Peterson, 1960. Pp. 223-248.
- SCHACHTER, S., NUTTIN, J., DEMONCHAU, CECILY, MAUCORPS, P. H., OSMER, D., DUIJKER, H., ROMMETVEIT, R., & ISRAEL, J. Cross-cultural experiments on threat and rejection. *Hum. Relat.*, 1954, 7, 403-440.
- SHERIF, M., & SHERIF, C. *Groups in harmony and tension*. New York: Harpers, 1943.
- WRIGHT, M. E. The influence of frustration on the social relations of children. *Charact. Pers.*, 1943, 12, 111-122.

(Received February 20, 1961)



## COGNITIVE RESTRUCTURING AND ATTITUDE CHANGE<sup>1</sup>

TIMOTHY C. BROCK

University of Pittsburgh

Most investigations of attitude change have been concerned with *evaluative* components of attitude, i.e., shifts in the direction and strength of affective responses to attitudinal stimuli. Asch (1952) called attention to the importance of *structural* properties, i.e., the patterning or organization of cognitive elements which comprise the attitude object. Although explicit concern with structural features was shown by Peak (1955), Smith, Bruner, and White (1956), Green (1954), and Rosenberg (1956), there have been few attitude change experiments dealing with modifications in both structural and evaluative aspects of attitudes. The present experiment examined the interaction of restructuring and revaluation as alternative and complementary modes of reduction of cognitive dissonance.

Festinger's (1957) theory of cognitive dissonance has implications for attitude change where a person carries out behavior contrary to his beliefs or opinions. Experiments on forced compliance show that the greater the cognitions supporting compliance to such discrepant behavior, the less the cognitive dissonance and consequent attitude change (as measured by evaluation) toward the initially disliked position (e.g., Cohen, Brehm, & Fleming, 1958; Festinger & Carlsmith, 1959). Two recent studies (Brehm & Cohen, 1959; Cohen, Terry, & Jones, 1959) have shown that cognitive dissonance decreases with decrease in feelings of personal volition or increase in force to comply. Consequently, more attitude change would be expected where individuals experience a high rather than a low degree of subjective choice in engaging in behavior contrary to their prior beliefs.

A second determinant of magnitude of forced compliance dissonance may be the ex-

tent to which the complier is confronted with the cognitions associated with his dissonant behavior. Such confrontation should have consequences for attitude change to the extent that the individual's attention and thinking are focused on the implications of his discrepant stand.

In the present study, choice and confrontation were varied to set up differing amounts of dissonance and consequent pressure toward attitude change. When a person feels that he has had little choice in carrying out discrepant behavior, little dissonance should be produced and confrontation with the implications of his action should bring heightened resistance to revaluation. However, under conditions where an individual feels it was entirely up to him whether or not he carried out the contrary behavior, dissonance should be produced; the more the confrontation, the more the dissonance and consequent attitude change in line with the discrepant behavior in order to reduce that dissonance. Thus, under high choice conditions, there will be more positive attitude change under high than under low confrontation; under low choice, there will be more positive attitude change under low than under high confrontation.

### *Attitude and Cognitive Structure*

Cognitive restructuring refers to changes in the organization of cognitive elements which comprise an attitude. Zajonc (1954, 1960) followed Lewin (e.g., 1951, pp. 83-84, 305-338) in designing empirical operations for measuring cognitive structure. Two of these measures were employed here: the number of categories used to group attitudinal elements (grouping); the number of relations among the elements (bonding). A problem for dissonance theory concerns the possibility that changes in grouping and bonding of attitudinal elements may interact with, or serve as well as, revaluation of those elements when dissonance has been aroused. A more gen-

<sup>1</sup> The present study is a portion of a dissertation presented to the faculty of the Graduate School at Yale University in candidacy for the degree of Doctor of Philosophy. The guidance and encouragement of A. R. Cohen and the criticisms of J. W. Brehm are gratefully acknowledged.

eral issue concerns the cognitive resources that may be evoked by a person when he experiences inconsistency between his beliefs and his behavior. Cognitive restructuring has been employed (Asch, 1952; Lewin, 1951) to describe characteristic responses under such circumstances. Changes in grouping and bonding of elements are here assumed to be modes of cognitive restructuring.

In the present study, subjects were given two measurement sequences after they had carried out behavior contrary to their beliefs. In the first, opportunity for restructuring of the beliefs preceded opportunity for revaluation of the beliefs; in the second sequence, revaluation came before restructuring. This counterbalancing was aimed at exploring the comparative effectiveness of dissonance reduction through one avenue (revaluation) with another possible avenue (cognitive restructuring). In the absence of theory concerning the effects of dissonance on restructuring, examination was made of: (a) the direct effects of choice, confrontation, and order of measurement, on restructuring; (b) the extent to which revaluation and restructuring were associated under different treatment combinations.

## METHOD

### General Design

After evaluative and structural measures were taken of the subjects' religious beliefs, they carried out contrary behavior by writing an essay opposed to these beliefs under conditions of high and low choice, and high and low confrontation. The essay topic—"Why I Would Like to Become a Catholic"—met the following criteria: (a) subjects were clearly opposed to the position they undertook to support, (b) they were as interested and ego involved as possible, (c) arguments and reasons for the essay position were known to the subjects. Following this attempt to arouse differential dissonance and consequent attitude change pressure, the same evaluative and structural measures were again administered in a balanced order.

### Subjects

Subjects were 183 Yale non-Catholic freshmen who were approached in their dorms. Each was asked, privately, to give his "present religious affiliation," and to say "how much he had considered giving it up in order to join another religious organization."

### Procedure

Premeasures of attitude and structure were administered under the guise of exploring a new sur-

vey method. The subject rated three eight-point Likert-type items: goodness of arguments and reasons heard for "becoming a Catholic"; degree of "sympathy and understanding" for "someone like myself who had become a Catholic"; extent of personal opposition to "becoming a Catholic."<sup>2</sup> In order to elicit whatever cognitions occurred to him when he thought about his "becoming a Catholic," the subject was given a packet of 16 alphabetized slips with the following printed instructions:

Complete the sentence, "For me becoming a Catholic would mean": in as many ways as you can. On each slip write a short, concise, phrase that completes this sentence. Write as many completions as easily come to your mind.

When the subject had written as many completions as he could, he indicated what *groups* he saw among the implications (slips) he had written by following these instructions:

Now lay out all the slips you have written in front of you. You will notice that some slips seem to belong with other slips. Indicate how groups naturally form among the slips you have written by writing the letters of the slips in each group next to the roman numerals at the left-hand side of this page. You may show as many groups as you want, but a slip may appear in only one group.

Finally, to determine what bonds (relations) the subject saw among his slips he was given these instructions:

Now consider each slip, separately, and ask yourself what other slips would have to be changed, modified, or excluded, if the slip you are considering were changed, modified, or excluded. For example, next to letter "A" write the letters of the other slips that would be affected if what slip "A" says were changed or no longer true.

Thus, *bonds* among cognitions were indicated by showing that changing, modifying, or rejecting one implication of becoming a Catholic would be accompanied by alterations in one or several other implications (Zajonc, 1954, 1960).

*Experimental manipulations.* A week later subjects were invited by phone to a room near their dorm for a "follow-up on the survey research." Of the 183 original subjects, 38 refused to participate again, all but 1 indicating pressures from study or social ac-

<sup>2</sup> The attitude items were suggested by a content analysis of essays on "becoming a Catholic" written by other Yale freshmen in a pilot phase of the present experiment. Three kinds of opposition to becoming a Catholic were displayed: personal opposition to this course of action; denunciation of reasons heard for becoming a Catholic; lack of sympathy for and failure to understand others who had converted. Three items were added to provide a baseline for estimating possible regression effects; they referred to the "elimination of intercollegiate athletics at Yale" and were otherwise identical with the "Catholic" items.



tivity. Only 1 subject refused because he felt offended at having to answer questions on religion. When subjects reported in small groups (2-4) to the experimental room, they were handed printed instructions guiding them (with the exception of the choice manipulation) through the remaining experimental events. Subjects read:

Forceful, creative, and compelling essays are needed on the following theme: "Why I Would Like to Become a Catholic." Even though this position may not be your own, do the best you can to write persuasive and original arguments.

At this point, in the High Choice condition, the experimenter said:

I would like to emphasize that even though you came over here tonight, there is no obligation at all to write this essay. If you don't want to write it you can get up and walk out if you wish. Is that clear?

Most subjects nodded. The experimenter turned to each subject individually and asked:

You know that its entirely up to you whether or not you write the essay? Are you sure?

All but seven subjects agreed and the manipulation was concluded after each had said "Yes" at least once. The loss of seven subjects who left the experiment showed that the option to write the essay was effectively communicated. (These 7 subjects and the 38 refusing to participate further in the experiment did not differ from the others on the premeasures of evaluation and structure. Therefore, it was felt that their attrition did not bias the results.) In the Low Choice condition, no such option was presented; after the subjects read the essay instructions, the experimenter said "please start writing now."

After 10 minutes, subjects in the High Confrontation condition were instructed to write out their essays again, ranking the sentences in terms of how "original and persuasive" they perceived their arguments to be. Here, the original material was copied over except that the sentences were reordered by the subject to show his assessment of their originality and persuasiveness (an emphasis on the "meaning" of the essay). In the Low Confrontation condition, subjects were instructed:

Copy your essays over using the number-of-syllables chart on the next page. List every word of your essay in the column appropriate to its number of syllables.

Under Low Confrontation, all the essay material was copied again but the emphasis was on grammar rather than meaning. Thus, the only difference between the High and Low Confrontation conditions was the extent to which the subject's thinking was focused on the implications of his reasons and arguments.

Following these manipulations, the premeasures were readministered. Approximately half the subjects rated the attitude scales and then were given a packet of slips to fill out and indicate groups and bonds. This condition, where the measure of evaluation

preceded the measure of structure, is labeled Evaluation-Structure. In the Structure-Evaluation condition, the measure of structure preceded the measure of evaluation.

A questionnaire assessed the effectiveness of the manipulations and the extent to which alternative modes of dissonance reduction, such as dissociation, had been employed. Finally, the experiment was explained and subjects were asked not to reveal its purposes for an appropriate length of time.

*Selection of subjects.* In order to assure that writing the essay was indeed dissonant, only those subjects were used who completed the attitude scale measuring personal opposition with opposed, very opposed, or completely opposed. A person who deviated on this attitude premeasure by more than a half-scale unit in the favorable direction from opposed was dropped from the entire study. Data from 11 subjects were omitted for this reason; thus, 127 subjects constituted the present sample.

## RESULTS

### *Effectiveness of the Experimental Manipulations*

In completing the final questionnaire, the subject indicated on eight-point scales his perceived degree of "obligation to write the essay" and the extent to which he felt "writing the essay was up to him." The High Choice subjects reported less obligation ( $t = 2.12$ ,  $p < .05$ ) and more option ( $t = 3.92$ ,  $p < .001$ ) than Low Choice subjects. Similar scales called the subject's attention to his re-writing of the essay: he was asked to report his degree of "awareness of the meaning of what" he wrote, and the extent to which he "deliberated upon the implications and consequences" of his "reasons and arguments." The High Confrontation subjects reported more awareness of meaning ( $t = 4.47$ ,  $p < .001$ ) and more deliberation ( $t = 6.24$ ,  $p < .001$ ) than the Low Confrontation subjects.

### *Effects of Choice, Confrontation, and Order of Measurement, on Attitude Change*

As anticipated, the three attitude scales measured somewhat different aspects of a person's feeling toward "becoming a Catholic": overall correlations of change scores were positive but not significantly different from zero. Since all three scales were considered meaningful and relevant, changes on each scale were combined by addition to provide an index of amount of revaluation of beliefs.

<sup>3</sup> All  $t$  tests in the present report are two-tailed unless otherwise noted.



TABLE 1  
MEAN AMOUNT OF ATTITUDE CHANGE BY EXPERIMENTAL CONDITION: ANALYSIS OF VARIANCE

Means							
	Structure-Evaluation				Evaluation-Structure		
	High Confrontation	Low Confrontation			High Confrontation	Low Confrontation	
High Choice	10.60(10)	7.69(13)	8.96	High Choice 7.53(53)	12.33(15)	0.53(15)	6.43
Low Choice	-14.46(16)	-2.11(19)	-7.74	Low Choice -0.43(74)	4.89(19)	7.30(20)	6.13
	-4.81	1.88	-1.12(58)		8.18	4.40	6.26(69)

Analysis of Variance <sup>a</sup>			
Source	df	MS	F
Choice	1	157.79	6.63**
Confrontation	1	0.00	<1
Order <sup>b</sup>	1	74.69	3.14*
Choice × Confrontation	1	108.56	4.56**
Choice × Order	1	139.46	5.85**
Confrontation × Order	1	37.66	1.58 <sup>c</sup>
Choice × Confrontation × Order	1	6.79	<1
Error	119	23.82	

Note.—Change on all three attitude scales combined. *N*s in parentheses. The more positive the mean, the greater the amount of reevaluation in the essay direction.

Key: Structure-Evaluation—Conditions in which opportunity for restructuring precedes opportunity for reevaluation. Evaluation-Structure—Opportunity for reevaluation precedes opportunity for restructuring.

<sup>a</sup> Because of unequal subclass frequencies, an approximation method by Walker and Lev (1953, p. 381) was employed.

<sup>b</sup> Order: prior opportunity for structural change vs. prior opportunity for evaluative change.

<sup>c</sup> Not significant.

\*  $p < .10$ .

\*\*  $p < .05$ .

following the writing of the essay. Support was found (Table 1) for the general expectation that more positive attitude change (in favor "of becoming a Catholic") would be obtained under High than under Low Choice conditions ( $p < .05$ ). In addition, the predicted interaction effect of choice and confrontation was obtained: under High Choice, more favorable reevaluation occurred under High than under Low Confrontation; under Low Choice, this effect was reversed. Order of measurement (Evaluation-Structure vs. Structure-Evaluation) did not affect the interaction: combining data for the order of measurement conditions yielded confirmation at an acceptable level ( $p < .05$ ).<sup>4</sup>

<sup>4</sup> Between-condition differences on each of the scales also supported the predicted interaction. Although it was obvious (after the fact) that more compelling confirmation ( $p < .001$ ) could have been obtained by ignoring changes on the first scale, subsequent mention of amount of reevaluation refers to combined changes as reported in Table 1.

Some possible relations between dissonance and cognitive restructuring were explored using the data in Table 1. First, if the restructuring operation (grouping and bonding implications of becoming a Catholic) reduced dissonance, less attitude change should be obtained when opportunity for restructuring preceded opportunity for reevaluation. Evidence in this direction was provided by an overall tendency for less positive attitude change in the Structure-Evaluation than in the Evaluation-Structure conditions:  $-1.12 < 6.26$  ( $t$  test  $p < .10$ ). Second, if the assumption was correct that restructuring, like reevaluation, reduced dissonance, then, where little dissonance was produced, one or the other of these avenues of reduction might suffice. Consequently, under Low Choice (low dissonance), less favorable reevaluation could be expected to occur when opportunity for it was preceded by opportunity for restructuring. Under Low Choice, less attitude change was obtained when opportunity for restructuring

turing came first:  $-7.74 < 6.13$  ( $t$  test  $p < .01$ ).

While the statistically large differences in Table 1 suggested that the restructuring operation reduced dissonance, one contrary difference was noted. Under High Choice, there was slightly more attitude change under the Structure-Evaluation conditions than under the Evaluation-Structure conditions:  $8.96 > 6.43$  (nonsignificant). This result, in conjunction with the reversal under Low Choice, accounted for the significant interaction between choice and order of measurement ( $p < .05$ ).

### *Effects of the Independent Variables on Amount of Restructuring*

The implications seen by each subject of his "becoming a Catholic" were analyzed in terms of the number of slips written, change in content from before to after writing the discrepant essay, change in the number of cognitive groups, and change in the number of cognitive bonds.<sup>5</sup> There were no differences between the eight experimental conditions in the sheer number of implications (slips) written. Both times the slips were written (before and after the essay) the overall average was about 7 slips, with a range from 3 to 16. There were similarly no differences in the extent to which new implications were indicated following the essay or in the extent to which pre-essay implications failed to be mentioned after the essay. Approximately 80% of the original implications were repeated in all experimental conditions. Persons very opposed to becoming a Catholic appeared to hold attitudes comprising elements that were stable and unchanging. Subjects in

<sup>5</sup> Typical implications were: "losing my friends," "giving up clear thinking," "not being able to practice birth control," "hurting my parents," "becoming part of a world organization," "being obliged to go to church every week," "feeling guilty more often than necessary," etc. The Protestants, Jews, agnostics, and atheists, who were approached by the experimenter, saw a variety of mainly unfavorable implications.

Where  $N$  = number of cognitive elements (slips written), change in groups signifies differences in groups/ $N$  from before to after essay writing; change in bonds signifies difference in bonds/ $N$  ( $N - 1$ ). Each denominator, of course, represents the possible number (of groups and bonds).

all conditions indicated about the same number of groups and bonds on the premeasure before the essay.

*Changes in cognitive grouping.* Change scores showed a general tendency toward fewer groups in all conditions except one. However, an analysis of variance yielded no significant outcomes and thus eliminated the possibility that the independent variables directly influenced regrouping. It was possible that whatever restructuring accomplished for the reduction of dissonance was achieved by decreasing rather than increasing the number of groups. The unavailability of control subjects (no essay writing) prevented exploration of this hypothesis.

*Change in cognitive bonding.* The mean amounts of change in number of bonds reported in Table 2 showed no main effects but the Choice  $\times$  Confrontation interaction was significant ( $p < .01$ ). Under High Choice, there was less decrease in bonds under High than under Low Confrontation; under Low Choice, less decrease under Low than under High Confrontation. These differences in amount of bond change exactly paralleled those for attitude change (see Choice  $\times$  Confrontation, Table 1). They suggested that relatively less decrease and/or increase in bonding reduced dissonance. "Less decrease in bonds," in spite of its descriptive accuracy, is conceptually unwieldy. Since specifying a bonds was somewhat laborious, doing it a second time (following the essay) probably evoked less interest and more fatigue factors, that may have accounted for an overall tendency to decrease bonds. Thus, the only significant outcome favored the possibility that dissonance was reduced by increasing relations among implications of "becoming a Catholic."

### *Correlational Results*

The extent to which change in groups and positive attitude change were associated appeared to be affected by the independent variations (Table 3). If decrease in groups facilitated favorable revaluation, these variables should be positively correlated under conditions where attitude change was relatively great. These conditions were, under High Choice, High Confrontation, and, under

TABLE 2  
MEAN AMOUNT OF CHANGE IN COGNITIVE BONDING BY EXPERIMENTAL CONDITION:  
ANALYSIS OF VARIANCE

Means							
	Structure-Evaluation				Evaluation-Structure		
	High Confrontation	Low Confrontation			High Confrontation	Low Confrontation	
High Choice	-5.80	-12.38	-9.52	High Choice -6.94	-0.07	-10.53	-4.97
Low Choice	-6.88	3.79	-1.09	Low Choice -5.51	-21.53	1.95	-9.49
	-6.45	-2.78	-4.43		-12.06	-3.11	-7.52
Analysis of Variance <sup>a</sup>							
Source		df	MS	F			
Choice		1	4.66	<1			
Confrontation		1	36.59	1.87 <sup>b</sup>			
Order		1	8.41	<1			
Choice × Confrontation		1	327.55	9.60***			
Choice × Order		1	73.92	2.17 <sup>b</sup>			
Confrontation × Order		1	11.47	<1			
Choice × Confrontation × Order		1	33.31	<1			
Error		119	34.09				

Note.—The more positive the mean, the greater the increase in number of cognitive bonds. Cell Ns are the same as in Table 1.  
<sup>a</sup> According to Walker and Lev (1953, p. 381).  
<sup>b</sup> Not significant.

\*\*\*  $p < .01$ .

Low Choice, Low Confrontation. Here, decrease in groups and positive attitude change were positively associated while, in other conditions, inverse relations occurred. Although only two of the four relevant comparisons in Table 3 were statistically significant, the pattern of differences in amount of correlation suggested that decrease in groups facilitated attitude change.

In a similar analysis of change in bonds (Table 4), substantial correlations between bond increase and positive attitude change were obtained only under conditions producing relatively greater attitude change. This outcome corroborated the direct effect of choice and confrontation on change in bonds. Hence, increasing relations among implications of "becoming a Catholic" may have facilitated favorable revaluation of anti-Catholic beliefs.

There were no significant differences between conditions in the extent to which change in groups and in bonds were associated. The overall correlation of these variables was only .12 ( $p > .20$ ).

### Alternative Avenues of Reduction and Artifacts

Opportunity was provided for two general modes of coping with dissonance: revaluation and restructuring. However, the results might be explained in terms of dissociation from the essay behavior, discounting the essay as an academic exercise, differential effort to write the best essay possible, perception of the experiment as meaningless. An analysis of the

TABLE 3  
RANK-ORDER CORRELATIONS BETWEEN DECREASE IN  
COGNITIVE GROUPS AND POSITIVE  
ATTITUDE CHANGE

	Structure-Evaluation		Evaluation-Structure	
	High Confrontation	Low Confrontation	High Confrontation	Low Confrontation
High Choice	+.28	-.31	+.22	-.15
Low Choice	-.39 <sup>a</sup>	+.40 <sup>a</sup>	-.43 <sup>b</sup>	+.18 <sup>b</sup>

Note.—Cell Ns are the same as in Table 1.  $p$  values computed for one-tailed  $t$  tests.  
<sup>a</sup> Difference between -.39 and +.40,  $p < .02$ .  
<sup>b</sup> Difference between -.43 and +.18,  $p < .04$ .



TABLE 4

RANK-ORDER CORRELATIONS BETWEEN INCREASE IN  
COGNITIVE BONDS AND POSITIVE  
ATTITUDE CHANGE

	Structure-Evaluation		Evaluation-Structure	
	High Con- frontation	Low Con- frontation	High Con- frontation	Low Con- frontation
High Choice	+ .52	+ .03	+ .45	- .03
Low Choice	- .37 <sup>a</sup>	+ .31 <sup>a</sup>	- .18 <sup>b</sup>	+ .36 <sup>b</sup>

Note.—Cell *N*s are the same as in Table 1. *p* values computed for one-tailed *t* tests.

<sup>a</sup> Difference between - .37 and + .31, *p* < .03.

<sup>b</sup> Difference between - .18 and + .36, *p* < .06.

attractiveness of these alternatives (as measured on the postquestionnaire) showed that they could not account for the results. Other eliminated explanations concerned potential artifacts: obtained attitude change was due to statistical regression; under some conditions more compelling essays were written; persons likely to be especially resistant to adopting a more favorable attitude toward becoming a Catholic, e.g., Jews and/or atheists, were disproportionately represented in the experimental subgroups. Detailed discussion of these points is found in Brock (1960).

### DISCUSSION

The attitude change results supported the theoretical deviations. Persons choosing to carry out behavior contrary to their opinions are strongly motivated to reduce the dissonance thus produced by revaluation of the opinions so that they are more favorable toward the discrepant behavior. This effect of choice was consistent with effects of this variable in other studies reviewed by Cohen (1960) and supported his summarization:

Where choice is varied, expectations from dissonance theory are fulfilled only under high-choice conditions; under low-choice conditions, straightforward motivational or resistance effects seem to account for the results (p. 306).

Under high dissonance (High Choice), opinion change in the direction of contrary behavior is greater to the extent that awareness of the meaning of the behavior and deliberation on its implications are heightened; when motivation to reduce opinion-behavior inconsistency is weak (Low Choice), confrontation with the behavior implications may

evoke "rebellion," strengthening of the original opinions, or, at least, resistance to revaluation in a direction consonant with the contrary behavior.

An exploration of the effects of dissonance on cognitive restructuring ruled out two possibilities: the restructuring operations *increased* dissonance; dissonance and restructuring were entirely unrelated. Change in cognitive groups was not directly influenced by the independent variations. The interaction effect of choice and confrontation on bond change did not allow an unequivocal interpretation. A consistent pattern of differences in the degree of association between each of the modes of restructuring and attitude change was noted. However, only half of the relevant comparisons were statistically significant. These results cannot be the basis for any definitive generalization about the relation between dissonance, revaluation, and restructuring. But they warrant the conclusion that, *when dissonance is aroused by carrying out contrary behavior, revaluation of cognitions referring to the original beliefs reduces dissonance more effectively than change in the way those beliefs are cognitively grouped and interrelated.*

If, following contrary behavior, beliefs were modified by increasing their interrelatedness and decreasing the number of groups among them, these changes might be interpreted in a number of ways. With respect to general determinants of dissonance (Festinger, 1957), it might be argued that the *salience* of the original beliefs or their *relevance* to the contrary behavior could be reduced by decreasing the number of categories required for grouping them. Salience might also be diminished by linking the beliefs with one another (increasing bonds) so that fewer beliefs are seen as autonomous or unrelated to the others.

Osgood and Tannenbaum's (1955) concept of incongruity is similar to dissonance. These authors might say that reduction of incongruity or "polarization along the evaluative dimension," was achieved by organizing "bad" implications of becoming a Catholic into a "tight" (many bonds) and "undifferentiated" (few groups) cluster.

Other interpretations stem from the *aversiveness* of contemplating negative implications of becoming a Catholic after creating

pro-Catholic propaganda. By increasing bonds and decreasing groups among these implications, differentiations among them could be lost; they could be more readily labeled in an omnibus fashion, consigned to "logic-tight compartments," and thus more easily suppressed. Too, the implications of becoming Catholic represented potential deprivations. Grouping them into fewer categories might enable them to be perceived as fewer losses; if one deprivation could be dismissed, increasing bonds among the implications might allow rejection of related deprivations as well. All of these interpretations overlap; further conceptual work and less ambiguous empirical returns are needed before a clear theoretical preference can be stated.

#### SUMMARY

The general objective was to investigate to what extent and under what conditions cognitive restructuring interacts with evaluative change to reduce cognitive dissonance. Dissonance was produced by varying the extent of choice and confrontation under which individuals engaged in discrepant behavior. The subjects were non-Catholic Yale freshmen who wrote persuasive essays in favor of "becoming a Catholic." After dissonance was created in this fashion, half of the subjects were first given the opportunity to reevaluate their former anti-Catholic beliefs and then given the opportunity to restructure these prior beliefs. The other half of the subjects were given these measures in reverse order. The reevaluation procedure consisted of administration of attitude scales relevant to "becoming a Catholic." The restructuring procedure consisted of instruments designed to elicit subjectively perceived implications of "becoming a Catholic," and the way in which these implications were grouped and interrelated.

The results showed an interaction between choice and confrontation in determining amount of attitude change. Persons who were given a choice to write or not to write the essay (high dissonance) evinced more change to the extent that they were confronted by the implications of what they had written (High Confrontation). Persons who were given no option to create pro-Catholic propaganda (low dissonance), evinced more

resistance to favorable change under High than under Low Confrontation.

The independent variables had no effects on cognitive grouping but cognitive interrelatedness, like revaluation, appeared to be an interactive function of choice and confrontation. Positive correlations were obtained between revaluation and increased homogeneity and interrelatedness under conditions producing relatively greater positive attitude change.

#### REFERENCES

- ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
- BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, **58**, 383-437.
- BROCK, T. C. Cognitive restructuring and attitude change. Unpublished doctoral dissertation, Yale University, 1960.
- COHEN, A. R. Attitudinal consequences of induced discrepancies between cognitions and behavior. *Publ. opin. Quart.*, 1960, **24**, 297-318.
- COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. soc. Psychol.*, 1958, **56**, 276-278.
- COHEN, A. R., TERRY, H., & JONES, C. Attitudinal effects of choice in exposure to counterpropaganda. *J. abnorm. soc. Psychol.*, 1959, **58**, 388-391.
- FESTINGER, L. *Theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.
- FESTINGER, L., & CARLSMITH, J. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203-210.
- GREEN, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
- LEWIN, K. *Field theory in social science*. New York: Harper, 1951.
- OSGOOD, C. E., & TANNENBAUM, P. H. The principle of incongruity in the prediction of attitude change. *Psychol. Rev.*, 1955, **62**, 42-55.
- PEAK, H. Attitude and motivation. In M. R. Jones (Ed.), *Nebraska symposium on Motivation: 1955*. Lincoln: Univer. Nebraska Press, 1955.
- ROSENBERG, M. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, **53**, 367-372.
- SMITH, M. B., BRUNER, J. S., & WHITE, R. W. *Opinions and personality*. New York: Wiley, 1956.
- WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
- ZAJONC, R. B. Cognitive structure and cognitive tuning. Unpublished doctoral dissertation, University of Michigan, 1954.
- ZAJONC, R. B. The process of cognitive tuning in communication. *J. abnorm. soc. Psychol.*, 1960, **61**, 159-167.

(Received March 1, 1961)



## AN INVESTIGATION OF CLINICAL JUDGMENT: A STUDY IN METHOD<sup>1</sup>

JOAN C. LEE<sup>2</sup> AND R. BENNETT TUCKER<sup>3</sup>

University of Kentucky

The clinical psychologist as diagnostician has provided a large share of the data available from the myriad studies of judgment already in the literature. Whether the clinician as subject is taken *in situ*, however, or observed under more controlled circumstances working with selected test protocols, two problems arise in understanding both the process and product of his judgments.

First, how are we to identify the cues actually used in making a judgment or prediction? Although most investigators provide the judge with certain specific data and ask him to base his prediction solely on that information, clinicians do, after all, acquire some biases and vagaries along the way. Rubin and Shontz (1960) for instance, present data showing that their judges were able to "describe" (using a *Q* sort technique) a paranoid schizophrenic patient on the basis of statistical data alone: date of birth, date of admission, education, occupation, etc. When given additional information in the form of a tape recording of an interview and test protocols, their descriptions changed very little. Although Rubin and Shontz term their judges' performances "a process of considered discrimination" one can never be certain whether the final diagnosis was primarily on the basis of age or high *F* + %. Even when we provide our clinician judge with the sparsest morsels of data in quantitative form, as Hoffman (1960) has done, we do not know what he adds to those data from his own frame of reference and ex-

perience nor precisely what contribution this makes to the total variance of his judgments.

A second problem lies in the choice of criterion for assessing the validity of a given judgment. Frequently, the judge's prediction is called right or wrong on the basis of a previous diagnosis by other clinicians who had access to more information regarding the patient. This is a somewhat questionable procedure in view of the evidence that the more information available to the judge, the less accurate his judgment (Gage, 1953; Giedt, 1955; Kostlan, 1954). Rubin and Shontz (1960) suggest that the more obvious the pathology, the more likely psychologists are to agree on the diagnosis. While most studies probably include some severe and perhaps "obviously" pathological cases in the sample to be judged, they must also include many borderline cases as well. The judge who disagrees with the criterion in a study of this type could well be right and the experimenter wrong.

The present paper describes two experiments in which quasiclinical judgment tasks were devised to study the judge's use of multiple cues. In both, naive judges were given tasks unique to them in order to minimize the influence of preconceived biases or individual level of competence on their performances. In each of the experiments the criterion can be clearly and objectively defined. It is suggested that this type of investigation circumvents the problems discussed above, yet still contributes to our understanding of the basic judgment process.

### EXPERIMENT I: THE COMBINING OF SEPARATE LEARNING EXPERIENCES IN PROBABLE CUE JUDGMENTS

Calvin and Curtin (1958) have described clinical judgment as "a synthesis of impressions." This type of judgment is probabilistic in that a decision is reached by weighing and

<sup>1</sup> This report is based upon two dissertations (Lee, 1960; Tucker, 1959) submitted to the Graduate School of the University of Kentucky in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Experiment I is part of a dissertation by R. Bennett Tucker under the direction of Betsy W. Estes. Experiment II reports data from a dissertation by Joan C. Lee with James S. Calvin as director of the study.

<sup>2</sup> Now at Howard University.

<sup>3</sup> Now at Veterans Administration Hospital, Lexington, Kentucky.



combining signs, or cues, which have varying probabilities. Thus, a judgment or diagnosis is reached by considering a number of relevant, probabilistic cues in the presence of other irrelevant and conflicting cues. In order to investigate this type of judgment, Calvin and Curtin devised a task in probability learning using series of cards varying in several respects: size, form, color, texture, dot-size and number of dots appearing on the card, and thickness. In a pilot study, three of these attributes were made relevant, probable cues and subjects were given the task of learning to identify X cards and Not-X cards. This appears to be analogous to the situation in which the clinician attempts to identify schizophrenic patients, for instance, from Rorschach determinants. The results of this preliminary study indicated that learning occurred, but that the subjects tended to overuse one of the three relevant cues in identifying X cards (i.e., to rely upon it beyond its degree of cue validity) and to underplay or totally neglect the other relevant cues. This type of one-cue learning persisted over many training trials despite the fact that subjects were informed of the correctness or incorrectness of their judgments. In the same task, but presented with two equally valid cues, another group of subjects showed the same tendency to overuse one preferred cue.

Because it has been shown that previous practice with simple concepts helps with the learning of more complex ones (Kendler &

TABLE 1  
PROPORTION OF THE THREE LEVELS OF COLOR AND  
SIZE CUES COMBINED AS INDICATORS  
OF X CARDS

Cue	Number of X cards	Number of Not-X cards	Percentage indicator of X
6-inch yellow	0	8	.00
6-inch red	2	6	.25
6-inch blue	4	4	.50
8-inch yellow	2	6	.25
8-inch red	4	4	.50
8-inch blue	6	2	.75
10-inch yellow	4	4	.50
10-inch red	6	2	.75
10-inch blue	8	0	1.00
Total	36	36	

TABLE 2

DIFFERENT TYPES OF CARDS HAVING THE SAME  
DEGREE OF VALIDITY AS INDICATORS OF X

Percentage indicator of X				
.00	.25	.50	.75	1.00
6-inch yellow	6-inch red 8-inch yellow	6-inch blue 8-inch red 10-inch yellow	10-inch red 8-inch blue	10-inch blue

Vineburg, 1954), it seemed likely that subjects who were required to learn to use a previously neglected cue might be more likely to make judgments on the basis of multiple, probable cues when they were again presented with the favored cue and the neglected cue together.

### Method

**Subjects.** Nine "bright normals," six males and three females, served as subjects. Although no formal test of intelligence was administered, subjects were sufficiently well known to the experimenter to enable him to make an estimate of their intelligence. These estimates based primarily upon verbal facility, amount of education, profession, and interests, were all of "high average" intelligence, or higher. Due to the length and difficulty of the task, factors such as motivation, cooperation, and available time were also taken into consideration in the choice of subjects.

**Materials.** Three sets of 72 training cards and one set of 72 test cards were constructed which varied in color, size, shape, texture, and size of dots appearing on them.

In this experiment, cue is defined as a discriminable attribute of the stimulus. A probable cue is one appearing less than 100% of the time as X. A relevant cue is an attribute which occurs more, or less, than 50% of the time relevant to X-ness, while an irrelevant cue occurs exactly 50% of the time relevant to X-ness and 50% of the time not relevant to X-ness.

### Condition I

For the 72 training cards of Condition I, color and size were selected as relevant cues. The proportion that various levels of these variables in combination was an indicator of X-ness is shown in Table 1. For instance, the 10-inch blue card was always X, while the 10-inch red card was called X by the experimenter 75% of the time. The 6-inch yellow card was never an X, and the 6-inch red card was called X only 25% of the time. Table 2 compares the different types of cards having the same degree of validity as indicators of X. It can be seen that, in gen-

eral, the red and blue cards are more likely to be called X than the yellow cards; and the larger the card, the more likely it is to be an X card.

Form, dot-size, and texture were made irrelevant cues; i.e., each level of these variables appeared 50% of the time as X and 50% of the time as Not-X.

These 72 cards were divided into two sets of 36 cards and called Training Series A and B. Seventy-two test cards for Condition I (also used in Conditions II and III) were constructed in exactly the same manner with color and size as the relevant cues, and dot-size, form, and texture as irrelevant cues. These 72 cards were then randomly divided into four series of 18 cards each and these series were called Test Series a, b, c, and d.

### Condition II

Two sets of 72 training cards were constructed for Condition II. In one of these sets, color was made the only cue relevant to X-ness. Size was held constant by making all of the cards 8 inches. The degree to which various colors were an indicator of X-ness was the same as in Condition I. Form, dot-size, and texture were kept as irrelevant cues. These 72 cards were randomly divided into two series of 36 cards each and called Training Series A and B.

In the other set of 72 cards, size was the only relevant cue. Color was held constant by making all cards in the series red. The degree to which various cards were indicators of X-ness was the same as in Condition I. These 72 cards were also divided into two series of 36 cards each and called Training Series A and B.

Since the same Test Series a, b, c, and d were used in Condition I and II, both color and size were cue variables in all test series.

### Condition III

The materials for Condition III were those used in Condition I and the procedure of Condition I was replicated save for the number of training and test trials.

**Procedure.** A minimum of 8 training and 8 test series for Conditions I and II was predetermined arbitrarily for all subjects. If, at the end of these series, the coefficient of cue utilization showed that the subject had not learned to use one or both of the relevant cues, the number of training and test series was increased to 12 or 16, i.e., until the coefficient of cue utilization differed significantly from zero ( $r = .34$ ). All subjects were given exactly 8 training and 8 test series in Condition III.

At the beginning session of each condition, the following instructions were read to the subject:

I am going to hand you some cards, one at a time. I want you to take each card and look at it. Some of the cards are called "X cards" and some are "Not-X." When I give you a card, I will name it. I will tell you whether it is an X or a Not-X. I want you to try to learn to identify, as well as you can, which cards are X and which are Not-X. You may hold a card as long as you like. As you

finish with the cards, please stack them, face down, to your right.

Training Series A was shown. In presenting each card, the experimenter said: "This is an X card"; or, "This is not an X card." After the completion of Training Series A, Test Series a was given, the following instructions being first read to the subject:

This is a test series. I want you to tell me whether you think each card is X or Not-X.

Training Series B was then presented in the same manner as Training Series A. Test Series b and subsequent test series were presented as Test Series a had been.

Instructions for Condition II and III were identical with those of Condition I. However, in Condition II those subjects who had shown a preference in Condition I for the color cue were given the training series in which size was the only relevant cue. Subjects who had relied primarily on the size cue in Condition I were now given the training series in which color was the only relevant cue. Thus, each subject in Condition II was trained on only one variable, color or size.

### Results

Cue validity is defined as the extent to which a cue, i.e., a given attribute of a card, is a valid indicator of X-ness, as computed by the point biserial coefficient of correlation. The validity coefficient for each of the relevant cues, color and size, as a probable indicator of X cards, was .41. In combination, these cues were probable indicators of X cards with a validity coefficient of .58. The extent to which a subject uses a given cue to categorize a card as X or Not-X may be estimated by correlating cue dimensions with the subject's judgments of X or Not-X. Coefficients of cue utilization were computed by combining two test series of 18 trials each, giving a total of 36 trials. With 33 *df*, coefficients of utilization differed with statistical significance from zero at the .05 level of confidence when  $r = .34$  or above.

Subjects may be said to use, overuse, underuse, or not use a relevant cue. For instance, if the coefficients of cue utilization for four subjects are .60, .22, .34, and .68, while actual cue validity is .41, the first subject is not using the cue. The second, whose coefficient is .22, is underusing the cue. The third subject is using the cue, and the fourth subject, whose  $r = .68$ , is overusing the cue since his reliance upon it exceeds the extent to which the cue is a valid indicator of X-ness.

Six of the nine subjects showed one-cue rather than multiple-cue learning in Condition I. These subjects tended to focus upon one relevant cue and to overuse this cue in making their judgments. Although several attempts to use both relevant cues, as well as incidents of switching from one cue to another as a basis for judgments, were noted, at the end of Condition I all of these subjects were overplaying the preferred cue to the almost total neglect of the other relevant cue. Two subjects showed a definite tendency to synthesize judgments by using both relevant cues in Condition I, but only one of the relevant cues was used consistently. One subject failed to learn the task in Condition I, and he was dropped from the experiment. Questioning revealed that he had failed to grasp the probabilistic nature of the task. Color happened to be the preferred cue for all subjects but one, who overused the size cue.

In the Test Series of Condition II, five subjects (3, 4, 5, 6, and 9) switched from their previously preferred relevant cue (color) and learned to make their judgments by using the size cue, the only cue reinforced for these subjects during the training series of Condition II. One subject (1) learned the previously neglected cue (size) and used it in making judgments but continued to overuse his previously preferred cue (color). Another subject (7) used both relevant cues early in the test series of Condition II. Then, the color cue, unreinforced in the training series, was abandoned and the size cue was greatly overused in terms of its cue validity. The only subject (8) who learned to operate on the basis of size in Condition I continued to use this cue in the test series of Condition II but did not use it at the level of statistical significance. The color cue, the only cue reinforced for this subject in the training series of Condition II, was learned and overused. Thus, all eight subjects learned in Condition II to make judgments on the basis of a previously neglected relevant cue.

There was considerable variability in the use of the two relevant cues in Condition III. Five subjects (1, 5, 6, 8, and 9) gave some evidence of having learned to combine their separate learning experiences and make multiple-cue judgments. Two other subjects (4 and

TABLE 3  
COEFFICIENTS OF CUE UTILIZATION FOR INDIVIDUAL SUBJECTS IN CONDITION III

Subjects	Cue	Test series			
		a	b	c	d
1	Color	.65	.55	.68	.42
	Size	.22	.28	.34	.70
	Combined	.61	.59	.72	.79
3	Color	.00	-.07	.07	.00
	Size	.68	.82	.76	.75
	Combined	.48	.53	.59	.53
4	Color	.76	.75	.70	.62
	Size	.15	.14	.21	.21
	Combined	.64	.63	.64	.58
5	Color	.35	.14	.48	.63
	Size	.21	.64	.48	.21
	Combined	.39	.55	.68	.59
6	Color	.34	.14	.48	.14
	Size	.00	.28	.41	.57
	Combined	.25	.30	.63	.50
7	Color	.00	.00	.28	.28
	Size	.22	.70	.64	.64
	Combined	.15	.49	.65	.65
8	Color	.68	.66	.84	.52
	Size	.27	.30	.07	.52
	Combined	.68	.68	.64	.73
9	Color	.27	.14	.62	.42
	Size	.75	.61	.48	.63
	Combined	.73	.53	.78	.74

7) used both relevant cues to making their judgments but tended to underplay one of these cues and to overuse the other. One subject demonstrated the ability to operate with either of the relevant cues separately but was unable to combine the two in Condition III. Coefficients of cue utilization in Condition III for the color and size cues separately and for the two combined are given in Table 3.

In this task, accuracy of judgment is defined as the percentage of times the subject's judgment coincided with the experimenter's calling a card X or Not-X in the training series. Subjects do not show the steadily rising curve typical of a learning task, perhaps because of the probabilistic nature of the task and the relatively few trials. Subjects continued to try out various cues, singly and in combinations, even after they became convinced that no cue would work perfectly. As



long as the training series continued, subjects would continue to test different hypotheses in an effort to improve their accuracy. Such hypotheses, when found not to pay off, were usually abandoned, and the subject would then make his judgments in the following test series by reverting to the color and size cues, singly or in combination.

Coefficients of cue utilization for all test series were combined for the group of eight subjects.  $z'$  values were obtained for all  $r$ 's (Edwards, 1954), and an average taken for each of the test series. The mean  $z'$  value for each test series was then converted back to the appropriate value of  $r$ . For subjects having more than the minimum of four  $r$ 's for any one condition, pairs of  $z'$  values were averaged. The resulting four coefficients of utilization for each variable in each of the three conditions are presented graphically in Figure 1. In Condition I, the group as a whole made increasingly greater use of the color cue, while neglecting the size cue. In Test Series d, the coefficients of cue utilization for color and size were .61 and .15, respectively. In Condition II, the group as a whole shifted to the size cue, with concomitant neglect of the color cue. Coefficients of cue utilization in Test Series d for size and color were .63 and .21,

respectively. It will be recalled that for seven of the subjects, size was the only cue being reinforced in the training series, and these subjects switched from the originally preferred color cue to the training cue as a basis for their judgments. In Condition III, both relevant cues were used at the .05 level by the group in judging the cards, coefficients of cue utilization for color and size being .40 and .55, respectively (see Figure 1).

### Discussion

At least two investigators (Goldberg, 1959; Hunt, 1959) have reported studies of clinical judgment in which nonprofessional judges were as successful as clinical psychologists in predicting from test data. In Goldberg's study trainees and staff psychologists did no better than a group of secretaries in diagnosing organic brain damage from the Bender Gestalt Test. All diagnosed above a chance level. One possible interpretation of Goldberg's study is that formal training may actually contribute little to the clinician's diagnostic skill.

The present data indicate that subjects can learn to combine relevant, probable cues and use these cues, in terms of their cue validity, in making judgments. It would appear that learning and practice with cues taken sepa-

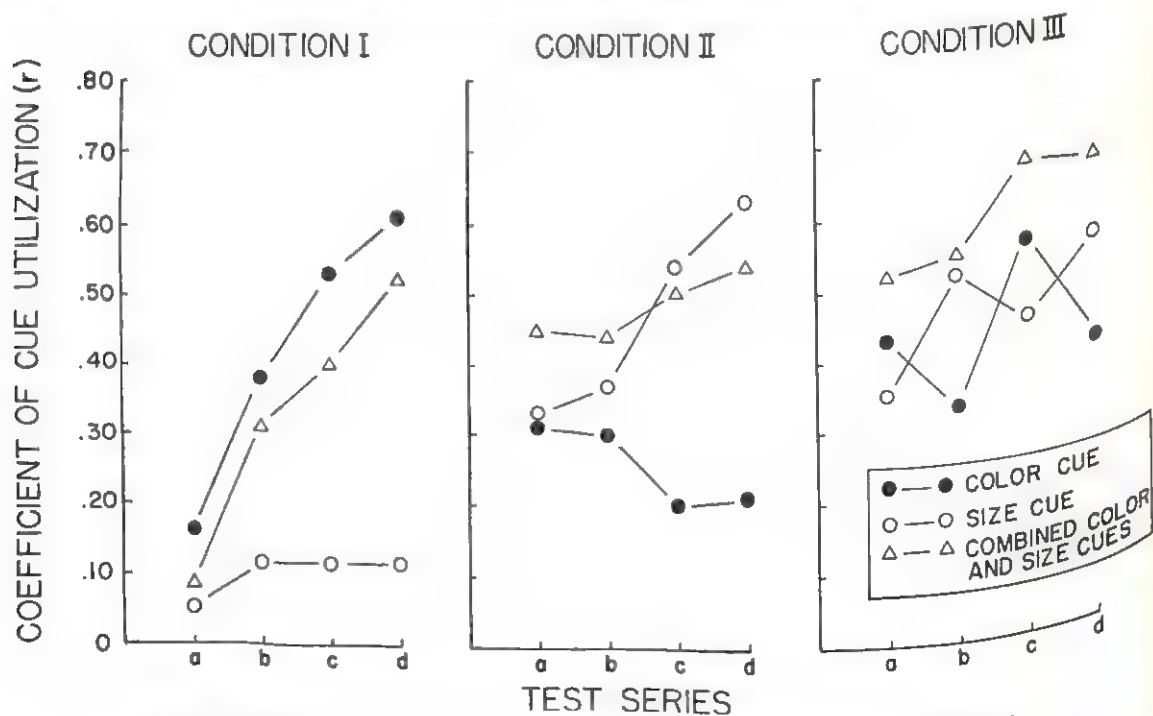


FIG. 1. Coefficients of cue utilization for color, size, and cues combined ( $N = 8$ ).

rately are important for the later use of them in combination. Holt's (1958) findings that his subjects were less successful than he had hoped in using carefully prepared manuals listing statistically validated cues might well be reversed if his judges had had some appropriate training in the use of the cues he provided them. Perhaps preliminary training trials requiring the judges to make predictions on the basis of each of the cues taken separately would have produced results similar to those we have reported.

Studies demonstrating that the judge *can* use a number of cues also suggest his limitations, as well. If the use of multiple cues can be facilitated by training, we must devise effective training methods and put them to use before we can give any final answer to the basic question, how much information can the judge process easily and efficiently?

#### EXPERIMENT II: CUE SYNTHESIS IN A QUASICLINICAL TASK

Psychologists develop their biases regarding test data long before they are asked to participate in an experimental investigation of clinical judgment. If we could persuade our clinician-subject to disregard his training insofar as it inhibits his good judgment, his predictions might represent maximum rather than typical performance. Since this is not easy to do, there is one alternative. "Naive" judges, i.e., subjects with no training in interpreting test data, can be placed in a situation analogous to that of the clinician who is asked to predict behavior from test scores.

An appropriate judgment task was conceived of as follows: one in which the judge would predict, from test scores, a type of behavior involving primarily those abilities which the tests measured. More specifically, a sort of "game" was constructed using abilities for which suitable measures could be found.

Thurstone (1938) has identified a number of different, relatively independent mental abilities which together make up what is called intelligence. These are measured by the Scientific Research Associates tests of Primary Mental Abilities (1949). The intermediate form of this test for ages 11-17 provides separate scores for spatial ability, rea-

soning, word fluency, verbal meaning, and number ability. The PMA was administered to a group of University of Kentucky freshmen. The low correlations between tests for this group of subjects indicated that the tests were measuring relatively independent mental abilities. On the basis of maximum scatter, three tests were selected for use in this study—number, reasoning, and space—and the game of Rocket was invented.

Rocket is a game for one player. The subject is asked to imagine himself a dispatcher in a space station somewhere above the planet Earth. His job is to check passengers and freight on every rocket leaving the station, and to see that each flight of three rockets has a squadron leader. He plays by going through three stacks of cards—passenger cards, freight cards, and rocket cards—and recording passengers cleared for each flight, freight OK'd, and squadron leaders on a tally sheet.

The correct identification of passengers for each rocket demands spatial ability. The subject is required to visualize a two-dimensional figure as it would appear if rotated in the same plane. All items for this part of the game were taken from the 1943 edition of the PMA and are of the same type as the space items appearing in the 1949 revision. In checking "stellar weights" on the freight cards, the subject must find all errors in a series of multiplication problems. The problems were taken from the 1943 edition of the PMA. All number problems in the 1949 revision require the subject to check the addition of columns of three two-digit figures. The subject identifies squadron leaders by correctly choosing the last in a series of symbols taken from Raven's (1958) Progressive Matrices Test. This requires reasoning ability, and is similar to the letter series found in the PMA.

The three PMA tests of spatial, reasoning, and number ability were administered to a total of 30 subjects, who were also asked to play Rocket for 20 minutes. The entire session required approximately an hour and a half. (Instructions and practice items for both tests and game require as much time as the actual testing.) Product-moment correlations between test and game for the subjects are

TABLE 4  
CORRELATIONS AND INTERCORRELATIONS BETWEEN  
PMA TESTS AND ROCKET

	Space	Number	Reasoning
Rocket score	.72	.32	.57
Space	—	.07	.23
Number	—	—	.03

reported in Table 4. The PMA abilities correlated with game performance to varying degrees. Correlations between PMA tests were negligible, except perhaps for the correlation between space and reasoning.

A multiple regression equation using the two best cues, space and reasoning, yielded a multiple correlation coefficient of .84 between actual and predicted scores for the 30 subjects. Optimal weighting of all three cues in a regression equation resulted in a multiple correlation of .88.

This set of 30 test protocols and criterion measures provided the material for the judgment situation. Given test scores reflecting the subject's number, space, and reasoning ability, the judge would be asked to predict how well that subject would do playing Rocket, a game requiring primarily those abilities (number, space, and reasoning) for which he had been tested.

Method

*Subjects.* The group of 30 judges included 9 males and 21 females, ranging in age from 18 to 35 years. They were introduced to the task by being told that they would be asked to make a series of predictions on the basis of test results in the same way that psychologists do. All judges were familiar with psychological tests and their purposes, but none of the judges had any training in testing.

*Procedure.* All judges followed the same procedure in making a total of 30 predictions in groups of 10. The judges began by familiarizing themselves with the tests and game. They read the PMA test instructions and worked the practice problems. The game was explained to them, and each judge worked through the first unit. The scoring system for tests and game was explained and maximum points on each indicated.

Each judge was then given a set of 10 test protocols and asked to try to rank the subjects according to their skill as Rocket players and predict how many of the total number of points possible they might earn. As soon as they completed their first series of predictions, they were given the scores actually earned by each of the 10 subjects in the

group and encouraged to compare these scores with their predictions.

Following their first series of 10 predictions, the PMA tests were administered to each judge under the same conditions as for the subjects. Then each judge played Rocket for 20 minutes. Test and game responses were scored immediately, and the judge was urged to compare the two. Then he was asked to predict for a second group of 10 subjects. As before, his predictions were corrected and errors called to his attention.

Before he made the final series of 10 predictions, each judge was told the relations between tests and game. Correlation coefficients were reported, and whether the judge seemed to understand this statistic or not, its meaning was explained to him. Judges were told, for instance, that although space was the "best" predictor (subjects tended to do as well on the game as they did on the space test), reasoning was also a good indicator and should be considered in predicting Rocket scores. It was suggested that number was a poor predictor and should be given little weight in making a decision.

After each set of 10 predictions, the judges were asked to describe how they had made their predictions. They were urged throughout to use any system they thought might maximize accuracy, and they were assured that there was no "correct" system that the experimenter expected them to discover.

Results

Hoffman (1960) and others (Hammond, 1955; Todd, 1954) have "described" the judge's method of combining cues with a multiple regression equation weighting each variable according to its contribution to the judge's predictions. This was done in the present study with equations fitted to each set of 10 predictions for all of the 30 judges. Table 5 summarizes the incidence of regression coefficients significantly different from zero for each variable in the three conditions.

Each group of 10 judgments may be considered one trial for the judge, with each

TABLE 5  
INCIDENCE OF SIGNIFICANT REGRESSION COEFFICIENTS  
( $p < .05$ )

Variable	Condition			Total
	I	II	III	
Space	26	30	28	84
Reasoning	14	13	13	40
Number	24	25	12	61
Total	64	68	53	185



TABLE 6  
NUMBER OF JUDGES ACCURATELY WEIGHTING  
VARIABLES IN EACH JUDGMENT SERIES

Condition	Number of judges
I	8*
II	8*
III	20**

\*  $p < .12$ .  
\*\*  $p < .001$ .

judge having three trials in all. In 84 trials, the regression coefficients for the space cue was significantly different from zero. Number score was used to a significant degree in 61 trials, but reasoning entered into the judges' predictions to the same degree in only 40 trials. Both space and number were often heavily weighted by the judges in making predictions in the first and second series of judgments. In the third series, however, when the judges had been advised that space and reasoning were the best indicators of game score, few judges continued to rely on the number cue. Equally few put more weight on reasoning, however, i.e., most judges used only the one best cue on their final series of judgments rather than the two best.

Considering the judges individually, we might ask how many of them achieved an accurate weighting of the three variables on any trial. Table 6 summarizes this information. If we disregard the absolute size of the weights and their significance and consider only their relative size (i.e.,  $S > R > N$ ), then 20 judges accurately weighted the test scores in making their final series of predictions. This would not be expected by chance.

The accuracy of this group of judges in predicting Rocket scores can be estimated by comparing the mean predicted scores for each subject with actually obtained scores. The product-moment correlation coefficient between mean predicted scores and actual scores in Condition III was .82. This was significant beyond the .001 level (from a significance test using Fisher's  $z$  transformation), as were the coefficients of .82 in Condition I and .79 in Condition II. The judges as a group were able to make accurate predictions of the subjects' performance in the game, but their judgments did not become more accurate with additional practice and experience with the criterion.

### Discussion

Because naive judges were used, it seems unlikely that preconceived ideas influenced their judgments to any significant degree. Their initial responses were made without training or special instructions as to how to proceed. We were interested in their approach to the problem—i.e., their analysis of

the test-game relation—as well as in their method of synthesizing the information given them. The judges' descriptions of the methods they used indicated that they perceived the task as one requiring the use of all the information. The ways in which they tried to do this were several, and most judges proposed two or three methods, modifying their system as they learned more about the criterion.

The data in Table 5 suggest that the judges got off to a good start, many of them correctly identifying at least one valid cue and actually using two on their first trial. When they were told what the best cues were, however, and urged to use both, they tended to abandon the number cue, as instructed; but instead of substituting the valid reasoning cue, most of them relied on the space cue alone in making the final series of predictions.

The explanation for the latter behavior (i.e., switching to one cue for the final series of predictions) may be obvious. This task was viewed by the judges as a difficult one. It is probably reasonable to suppose that whatever their method, they were eager to find a simple way of arriving at an accurate prediction. Of course, knowing that the space score alone was the best predictor, ranking by space alone was the simplest method, and the experimenter had guaranteed them a fair degree of accuracy using the procedure.

Considering only the judges' performance on Conditions I and II, however, it appears that many of them did put at least two cues together and come up with a fairly accurate prediction.

### CONCLUSIONS AND SUMMARY

Experimental investigations of clinical judgment encounter certain methodological diffi-

culties. The present study describes two quasi-clinical situations in which naive judges were asked to make judgments on the basis of multiple cues. In both of these experimental situations, the criterion for accuracy of judgment could be clearly and categorically defined. It was also possible to identify the basis for the individual judge's prediction.

In this context, judges proved able to handle at least two relevant cues in making a prediction. Experiment I suggested that prior training with the individual cues taken singly enhances the judge's ability to use both together when the opportunity is presented. Subjects in Experiment II showed a tendency to rely on a single cue when they were assured of a fair degree of accuracy in so doing.

In the study by Goldberg (1959) described above, an "expert" was called in and matched against both clinicians and secretaries in diagnosing brain damage from the Bender Gestalt. Where the others had spent from 45 minutes to 1 hour on the task, this man took about 20 hours, studying the records and making his predictions. He was able to better the record of the clinical trainees by correctly identifying 83% of the records as organic or nonorganic. (The clinicians had averaged 70% correct with the best of them getting 77% correct.) We might ask whether this slight advantage was worth the extra hours. Making judgments on the basis of multiple cues is hard work, and, in practice, the judge may find it expedient to sacrifice the accuracy to be achieved only by long and laborious pondering over test data. Certainly judges will be motivated to work so hard only if it can be demonstrated that their effort will result in some significant increase in accuracy.

Neither of the studies reported here was designed primarily to investigate accuracy of judgment. Actually, both groups of judges were fairly accurate, but it is not possible to show that there was much improvement in performance when judges based their predictions on two or more cues rather than one. What these studies do indicate, however, is that certain conditions facilitate multiple-cue

prediction. It is suggested that this type of study can be used to investigate accuracy of judgment as a function of amount of information processed by the judge, as well as to explore the limits of the human being in this respect.

## REFERENCES

- CALVIN, J. S., & CURTIN, MARY E. Learning of a probable-cue judgment. Paper read at Southeastern Psychological Association, Atlanta, April 1958.
- EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.
- GAGE, N. L. Explorations in the understanding of others. *Educ. psychol. Measmt.*, 1953, 13, 14-26.
- GIEDT, F. H. Comparison of visual, content, and auditory cues in interviewing. *J. consult. Psychol.*, 1955, 19, 407-416.
- GOLDBERG, L. R. The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender Gestalt Test. *J. consult. Psychol.*, 1959, 23, 25-33.
- HAMMOND, K. R. Probabilistic functioning and the clinical method. *Psychol. Rev.*, 1955, 62, 255-262.
- HOFFMAN, P. J. The paramorphic representation of clinical judgment. *Psychol. Bull.*, 1960, 57, 116-131.
- HOLT, R. R. Clinical and statistical prediction: A reformulation and some new data. *J. abnorm. soc. Psychol.*, 1958, 56, 1-12.
- HUNT, W. A. An actuarial approach to clinical judgment. In B. M. Bass & I. A. Berg (Eds.), *Objective approaches to personality assessment*. Princeton: Van Nostrand, 1959. Pp. 169-191.
- KENDLER, H. H., & VINEBERG, R. The acquisition of compound concepts as a function of previous training. *J. exp. Psychol.*, 1954, 48, 252-258.
- KOSTLAN, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.*, 1954, 18, 83-88.
- LEE, JOAN C. Analysis and synthesis in judgment. Unpublished doctoral dissertation, University of Kentucky, 1960.
- RUBIN, M., & SHONTZ, F. C. Diagnostic prototypes and diagnostic processes of clinical psychologists. *J. consult. Psychol.*, 1960, 24, 234-239.
- TODD, F. J. A methodological study of clinical judgment. Unpublished doctoral dissertation, University of Colorado, 1954.
- THURSTONE, L. L. Primary mental abilities. *Psychometr. Monogr.*, 1938, No. 1.
- TUCKER, R. B. The combining of separate learning experiences in probable cue judgments. Unpublished doctoral dissertation, University of Kentucky, 1959.

(Received March 1, 1961)

## FREQUENCY OF TREATMENT AND CHANGE IN PSYCHOTHERAPY

MAURICE LORR, DOUGLAS M. McNAIR, WILLIAM W. MICHAUX, AND ALLEN RASKIN<sup>1</sup>

*Veterans Benefits Office, Washington, D. C.*

Currently most psychiatric outpatients who are seen for psychotherapy in public clinics are scheduled for treatment once or twice a week for about an hour on each occasion (Feldman, Lorr, & Russell, 1958; Hollingshead & Redlich, 1958). These treatment frequencies are based primarily on tradition and clinical experience. There is very little systematic research information concerning the efficacy of this or other patterns of interview frequency. Actually therapeutic progress could have linear or curvilinear relations with the number of treatments received over fixed intervals. Ordinarily, it is lack of therapist time and monetary considerations that tend to limit the number of contacts per week. However, one expectation, widely held, is that therapeutic gain increases in rough proportion to the number of treatments. This belief receives some support in reports by Seeman (1954); Myers and Auld (1955); Imber, Frank, Nash, Stone, and Gliedman (1957); and Feldman et al. (1958). On the other hand, Clara Thompson (1950) argues that it is duration rather than frequency of contact that is the crucial variable.

In many respects treatment frequency is like the dosage level of a drug. Like dosage level, frequency of treatment should have important relations to treatment effects, to patient variables such as severity and nature of the disorder, and to treatment variables such as therapist characteristics and the method of treatment. The present study represents an

effort to explore a few of these relationships. Its major purpose is to test the hypothesis that therapeutic gains resulting from individual psychotherapy increase with the number of treatment interviews received over fixed time intervals.

### *Hypothesized Therapeutic Outcomes*

The dependent variables assessed in the study were selected on the basis of a survey of the literature as to the kinds of behavior purported to change as a function of psychotherapy. The viewpoint taken was that, as a result of psychotherapy, patients might be expected to report feeling more comfortable; report fewer or less disturbing physical or psychic complaints; function more effectively and comfortably interpersonally; and accept themselves more realistically as individuals.

The changes hypothesized may be particularized as follows: the greater the number of psychotherapeutic interviews within a specified period, the greater the:

1. Reduction in manifest anxiety
2. Reduction in the number and severity of patient complaints and problems
3. Increase in ego strength
4. Increase in self-acceptance
5. Increase in sociability
6. Reduction in hostility towards others
7. Increase in independent behavior
8. Number of specific positive behavioral changes reported by the therapist
9. Level of self-awareness or understanding

### METHOD

#### *Study Plan*

The study design called for the random assignment of patients at each of seven mental hygiene clinics, to one of three different treatment schedules—twice weekly, once weekly, and once biweekly. The largest participating clinic randomly assigned patients to all three treatment frequencies. Two clinics assigned patients to once weekly and biweekly schedules. The remaining four clinics distributed patients randomly

<sup>1</sup> In collaboration with Richard W. Boyd, Louis Rutledge, Henry H. Weiss, Bernard B. Yormak, Philip W. Morse, John E. Tucker, and Harold J. Fine.

We are pleased to acknowledge the assistance of Eli A. Rubinstein and Martin M. Katz, now with the National Institutes of Health, in the planning and instrument development stages of the study. Special acknowledgment is due Elizabeth E. Turk, Velma Boone, and Barbara S. Gershberg for invaluable assistance in the processing and analysis of data and in the preparation of the manuscript.



to twice and once weekly treatment schedules. The contributing Veterans Administration clinics were located in Albany, Boston, Bridgeport, Buffalo, Chicago, Denver, and Hartford.

Because of the high incidence of dropouts it seemed imperative to re-evaluate the groups at the earliest point at which treatment effects might be manifested. Four months was chosen by the participating clinics as a period within which therapeutic gains could be expected for the types of patients studied. A second re-evaluation was scheduled after 8 months of psychotherapy in order to determine what additional gains would be demonstrated. Each patient selected for the study was interviewed and tested just before initiation of treatment and again at the end of 16 weeks and 32 weeks of psychotherapy. For the initial examination, the patient was seen successively by the intake social worker, the intake psychiatrist, and by the psychologist who administered the test battery which ordinarily required about an hour and a half. He was also rated by his therapist immediately after his first therapy interview. For the 16-week re-evaluation each patient was reinterviewed by the social worker, retested by the psychologist on an abbreviated test battery, and rerated by his therapist. The 32-week re-evaluations were identical with the 16-week re-evaluations except that the social worker's interview was discontinued for administrative reasons.

### *Sample*

The sample in each clinic was confined to male veterans with service connected psychiatric disabilities who were less than 51 years of age and without any present indication of brain injury. None had received 3 months or more of intensive psychotherapy within 90 days of inclusion in the study. All were acceptable to the clinics for "intensive" individual psychotherapy as subsequently defined.

The typical patient, of the 133 included, was 37 years of age although the range was from 21 to 51. Approximately half of all patients were high school or college graduates. Of those employed the average annual earnings was \$3,500. Seventy-five percent of the group were employed at the time of inclusion in the study. The typical patient's illness was rated as "moderate" on a four-point severity scale based on composite ratings obtained from psychiatrist, social worker, and therapist.

### *Treatment and the Therapists*

Most psychotherapists had a psychoanalytic orientation to treatment although modified Rogerian and Sullivanian approaches were also represented. The term "intensive psychotherapy" was defined in terms of interviews of about an hour in duration and treatment directed towards reorienting or changing the patient, assisting him in modifying his personal adjustment patterns, and aiding him in making more constructive use of his assets. Patients whose treatment would necessarily be directed at keeping them out of a hospital or at maintaining

them at present levels of adjustment were to be excluded from the study.

A therapist was defined as any staff member or trainee whom the clinic regarded as competent to conduct intensive psychotherapy. Seventy-five different therapists participated in the study. Of these, 64 were staff members and 11 residents or trainees. Twenty-three therapists were psychiatrists, 23 were social workers, and 29 were clinical psychologists. Their years of experience as therapists averaged 4.5 years.

### *Patient Measures*

The 10 patient criteria used to evaluate the changes hypothesized were the following:

1. A 50-item modified version of the Manifest Anxiety (MA) scale (Taylor, 1953) was used to test the hypothesis concerning manifest anxiety. Reports show that the scale correlates well with the clinical concept of anxiety.
2. The Symptom Checklist was designed to evaluate the hypothesis concerning the number of patient complaints. Included were 20 of the most common complaints reported by clinic patients.
3. The 56-item Ego Strength scale (Barron, 1953), also taken from the MMPI, was used to test the hypothesis concerning ego strength. A few items concerning religion, potentially offensive, were omitted. Barron considers the score to be an estimate of adaptability and personal resourcefulness; a measure of general capacity for personal integration.
4. A 15-item Sociability measure was taken from the Guilford-Zimmerman Temperament Survey (Guilford & Zimmerman, 1949). A high score purports to be a measure of liking for people and ease in making social contacts with others.
5. The 15-item Friendliness scale, also taken from the Guilford-Zimmerman Temperament Survey, was used to assess hostility. A low score is indicative of hostility, belligerence, and suspicion of others. A high score indicates respect for others and tolerance of hostile acts by others.
6. The Self-Rating scale consists of 16 five-point linear self-descriptive scales concerned with self-satisfaction. It was included as a measure of self-acceptance and self-satisfaction. The lower the score the greater the dissatisfaction with the self.
7. The Interpersonal Checklist (ICL) (La Forge & Suczek, 1955; Leary, 1957) constitutes the source for the ICL criteria—assertiveness, cooperativeness, hostility, and dependence. The Assertiveness score was based on 17 ICL adjectives that a group of psychologists judged "therapeutically desirable." The score is a measure of assertiveness, self-reliance, competitiveness, and firmness with others.
8. Another 12 ICL adjectives, also judged to describe therapeutically desirable behavior, formed a Cooperativeness-Responsibility scale. A high score describes a relatively modest, cooperative, and responsible individual.
9. The Autocratic-Hostile scale is composed of 29 adjectives in the ICL judged to describe "thera-

peutically undesirable" behavior such as aggressiveness, exploitiveness, resentment, and suspiciousness.

10. The Dependent-Docile scale consists of 30 adjectives and phrases taken from the ICL which are descriptive of such characteristics as passivity, self-effacement, overdependency, excessive generosity, and undue deference. All items were judged to be therapeutically undesirable.

In addition to the criteria of change, a number of test measures were included as possible predictors of change. Twenty-five statements from the F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) were used as a measure of authoritarianism. Low scores indicate rigid adherence to conventional values, condemnation of persons who violate conventional standards, and generalized hostility. A measure of interest in thinking, observing oneself and observing others, called Reflectiveness, was obtained from the Guilford-Zimmerman scales. It was conjectured that patients responsive to psychotherapy are more introspective than those who are not. An adapted 50-item Behavior Disturbance Inventory was included as a measure of psychopathic tendencies (Appelzweig, Dibner, & Osbourne, 1958). It purports to identify individuals who are restless, aggressive, nomadic, and hostile to authority. A multiple-choice vocabulary test and a word fluency measure were also included as likely predictors of therapeutic response.

Coefficients of internal consistency were computed for each of the 15 patient and therapist criteria. Their median value was .80 and the range was from .68 to .89. To secure indices of test stability, pretreatment scores were correlated with 4-month and 8-month scores. The median correlation between the pretreatment versus 8-month scores was .78. These coefficients imply that the criteria are fairly homogeneous and surprisingly stable.

### *Therapist Measures*

Each therapist described his patient on a number of measures which were used to assess change from the therapist's viewpoint.

1. A Severity of Illness score was derived from ratings on seven four-point scales assessing characteristics related to degrees of severity of illness (Lorr, Holsopple, & Turk, 1956). Ratings were obtained immediately after the first interview and again at the time of each re-evaluation.

2. Each therapist was asked to describe his patient on the ICL. The degree of agreement between the therapist's description of the patient and the patient's self-description was then summarized by means of a concomitance index called the Patient-Therapist ICL Index (P-T ICL). It was conjectured that the extent of agreement would increase with progress in therapy and reflect the patient's understanding of himself.

3. The Change Inventory was completed by the therapist at the time of each re-evaluation (McNair & Lorr, 1960a, 1960b). The Inventory consisted of 92 statements marked true or not true, which are descriptive of specific changes frequently observed in

psychotherapy patients. An Interview Relationship (IR) score is based on 24 statements descriptive of the patient's participation and resistance during the interview. A second score of Interpersonal Changes (IC) is based on 42 specific changes in interpersonal relations judged to be therapeutically desirable. The third measure of Symptom Reduction (SR) is based on 26 observed reductions in the frequency or severity of symptoms and problems exhibited.

Therapists also rated the patient on a number of variables that held promise as predictors of response or which were needed as control variables. Each therapist rated his patient on four-point scales of Liking for the Patient and Interest in the Patient's Problem. He also rated the patient's degree of Motivation for Treatment. Finally, the therapist indicated the appropriateness of the randomly assigned treatment frequency for the patient's condition on a Suitability of Treatment Frequency scale.

### *Social Worker Measures*

The social worker interviewed the patient at the time of intake and again at the end of 16 weeks of treatment. Seven scales were concerned with Employment Adjustment, four scales described Social and Community Adjustment while three scales dealt with Family Life Adjustment. Only social adjustment and severity of illness ratings were used in the analysis.

### *Method of Analysis*

The statistical model for evaluating the effect of increased amounts of psychotherapy on the various criteria was analysis of covariance within a simple randomized design. Final criterion mean scores were adjusted for initial status on the criterion being analyzed as well as for the effects of other control variables as seemed necessary. In each instance the correlations between initial criterion scores, control variables, and final criterion scores were examined. As required the final criterion scores were adjusted by covariance both for initial criterion scores as well as for the background variables. The effect of adjustment was to provide statistical equality of the treatment groups prior to treatment and to reduce the size of the error term in the *F* test in evaluating final adjusted means. The patient control variables considered were: age, highest grade completed, annual earnings, employment status, vocabulary score, F Scale (authoritarianism) score, Behavior Disturbance scale score, and Word Fluency test score. Observer control variables available were: length of server control variables available were: length of therapist experience, rated competence of therapist, rated liking and interest of therapist for patient, and rated suitability of treatment frequency.

### *RESULTS*

The results for the 4- and 8-month assessment periods are presented sequentially. For each period the differential effects of treatment frequency are first described and then



TABLE 1

ADJUSTED TREATMENT GROUP MEANS AFTER FOUR MONTHS OF PSYCHOTHERAPY  
(Initial scores controlled)

Criterion	Treatment frequency			<i>F</i> <sup>a</sup>
	Twice weekly	Once weekly	Biweekly	
Patient measure:				
Manifest Anxiety	28.5	29.2	28.9	0.14
Ego Strength	30.3	32.2	31.5	0.34
Symptom Checklist	8.0	7.5	7.6	0.43
Self-Rating	52.8	53.8	52.3	0.13
Sociability	7.4	7.6	7.6	0.08
Friendliness	8.2	8.6	9.1	1.16
Therapist measure:				
Severity of Illness	20.6	20.7	21.0	0.12
IR	9.2	9.3	7.3	1.63 <sup>b</sup>
IC	7.2	7.2	5.0	1.90 <sup>b</sup>
SR	5.3	5.7	3.6	2.66 <sup>b</sup>
Other measure:				
SW Severity of Illness	20.0	19.7	19.3	0.63
Social Adjustment	29.5	29.2	27.6	0.53
P-T ICL	0.41	0.41	0.37	0.95
<i>N</i>	45	62	26	

<sup>a</sup> *F*<sub>.05</sub> (*df* = 2/129) = 3.06.

<sup>b</sup> Actual 4-month means are presented and tested as there were no initial scores on this variable.

the treatment gains. Whenever significant differences between treatment frequency groups could not be demonstrated, tests were made to identify any changes from pretreatment status. In order to test for total treatment gains, all treatment groups were pooled and tests were made on the combined group scores.

#### Treatment Frequency: Four Months

During the first 4 months of therapy, the twice weekly frequency group actually attended an average of 25.5 (*s* = 3.8) therapy sessions; the once weekly group, an average of 14.5 (*s* = 2.3) sessions; and the biweekly group, an average of 8.6 (*s* = 1.6) interviews. The three treatment groups did not differ significantly when treatment began on any of the criteria or on predictor or background variables. Table 1 presents the results of tests for treatment effects on 13 criteria of response to psychotherapy. The influence of initial scores on the 4-month scores is controlled by the method of analysis of covariance except in the case of the Therapist Change Inventories. Since there were no initial scores on the three Change Inventories, differences were tested by one-way analysis of variance.

None of the *F* tests on the criteria is significant at the .05 level. In addition to the criteria listed in Table 1, the four ICL scores also showed no differences at 4 months. With the exception of the Therapist Change Inventories, there is no evidence of a trend for the biweekly group to show less change than the two groups receiving more frequent interviews. The *F* test for SR Changes approaches significance (.10 > *p* > .05), and this result affords the only suggestion of a difference in response to psychotherapy as a function of interview frequency. Three sources of information—patient, therapist, and an independent social worker—apparently concur in observing no differences ascribable to treatment frequency. Thus the major hypothesis receives no support for the patients studied over a 4-month interval.

Was there a differential dropout rate by treatment frequency group prior to the 4-month evaluation? If so, and if patients with different characteristics dropped out of the different frequency groups, a serious bias could have been introduced into the comparisons. A chi square test indicated that there is no significant association between treatment frequency and dropping out or remaining. Additional analyses (not presented here) indicate that the terminators on each treatment

TABLE 2  
CHANGES OVER FOUR-MONTH PERIOD FOR  
COMBINED GROUPS

Criterion	Means		<i>t</i> <sup>a</sup>
	Initial	4 month	
Patient measure:			
Manifest Anxiety	29.4	28.9	0.83
Ego Strength	31.5	32.1	1.36
Symptom Checklist	7.9	7.7	1.01
Self-Rating	52.2	53.5	1.39
Sociability	7.5	7.5	0.20
Friendliness	8.8	8.6	1.15
Therapist measure:			
Severity of Illness	21.6	20.7	2.61**
Other measure:			
SW Severity of Illness	20.9	19.8	2.20*
Social Adjustment	30.4	29.0	1.67*
P-T ICL	35.4	40.0	3.74**
<i>N</i>	133		

<sup>a</sup> *t*<sub>.05</sub> (one-tailed) = 1.64.

\* *p* < .05.

\*\* *p* < .01.



schedule also do not differ significantly in initial status on the criterion or predictor measures or on background variables such as age and education.

#### *Treatment Gains: Four Months*

Table 2 presents initial and 4-month means and *t* tests on the 10 criteria which are administered on both occasions. There are no significant changes on the patient measures. However, both therapists and social workers independently observe a significant decrease in Severity of Illness of the study sample. The Patient-Therapist ICL Concomitance Index also increases significantly, i.e., patient and therapist are more similar in describing the patient on an adjective checklist after 4 months of therapy than after one interview. The ICL change appears due to therapists simply checking more adjectives in describing a patient after knowing him 4 months rather than to any shift in the content of the patient's self-descriptions. Thus, after 4 months of psychotherapy, the patients in the study report no significant changes on the patient criteria. However, therapists and social workers observe some favorable changes not associated with frequency of interviews.

#### *Treatment Frequency: Eight Months*

The 8-month analysis is based on 58 patients who remained in the study on the original randomly assigned treatment frequency schedules. There were 16 patients in the twice weekly group with a mean of 50.8 ( $s = 6.4$ ) therapy sessions. Thirty patients in the once weekly group had a mean of 27.7 ( $s = 4.1$ ) sessions, and 12 patients in the bi-weekly group had a mean of 14.4 ( $s = 2.8$ ) hours of treatment. Before starting treatment the groups did not differ significantly on criterion, predictor, or background variables. However, before considering the results of treatment with these 58 cases, two questions about the 8-month sample should be answered. Is loss of patients from the study systematic by treatment frequency? Is the 8-month study sample comparable to the group analyzed at the 4-month evaluation?

A chi square test of the relation between treatment frequency and remaining or dropping out of the study prior to the 8-month

evaluation was not significant although a slightly higher proportion of the twice weekly group did drop out during this time interval. The 75 cases lost from the study included 57 who had completed or stopped treatment and 18 others who remained in psychotherapy but were dropped from the study because their treatment frequency was altered for one contingency or another (e.g., employment interfered, therapist judged it imperative to change the frequency of interviews).

The pretreatment status of patients remaining in the study at 8 months at the assigned treatment schedules differed significantly in certain respects from the patients who dropped out. The differences suggest principally that the 8-month sample was more disturbed when they began therapy; *MA* scores were higher and self-descriptions were less favorable. The 8-month sample was also an average of 2 years older, but this difference was not quite significant. Therapists observed significantly less *IC + SR Change* in the 8-month sample during the first 4 months of therapy. However, an analysis of patient self-report measures indicated that the remainers and dropouts did not differ significantly in amount of change over 4 months. Thus patient reports did not corroborate those of the therapists.

Table 3 presents the results of a comparison of treatment frequency groups, after 8 months of treatment, on 15 criteria of response to psychotherapy. The *IC* and *SR* changes correlated so highly that they were combined into a single measure. None of the differences between frequency groups are significant at the .05 level, nor is there much suggestion of any observable trend in the hypothesized direction. Patient and therapist measures corroborate each other in indicating no differences between groups over the 8-month treatment period. Only the difference between groups on the patient-therapist *ICL* approaches significance. The difference suggests that the more frequently patients and therapists see each other, the greater their agreement in describing the patient. Thus once more the research hypothesis that treatment effects will be greater with more frequent treatments is not supported after 8 months of psychotherapy.

TABLE 3

ADJUSTED TREATMENT GROUP MEANS AFTER EIGHT MONTHS OF PSYCHOTHERAPY  
(Initial scores controlled)

Criterion	Treatment frequency			F <sup>a</sup>
	Twice weekly	Once weekly	Biweekly	
Patient measure:				
Manifest Anxiety	29.1	30.4	30.5	0.22
Ego Strength	32.2	32.8	31.9	0.16
Symptom Checklist	7.6	7.4	7.3	0.05
Self-Rating	46.2	44.8	47.4	0.53
Sociability	6.9	6.1	6.4	0.63
Friendliness	8.2	8.6	9.9	1.90
ICL: Assertiveness	9.7	9.2	9.1	0.23
ICL: Cooperativeness	7.4	8.0	8.9	2.16
ICL: Hostility	12.6	11.8	11.6	0.38
ICL: Dependence	15.8	14.7	14.9	0.45
Therapist measure:				
Severity of Illness	18.9	19.9	21.0	1.35
IR	8.9	9.2	7.6	0.62 <sup>b</sup>
IC + SR	13.8	11.5	11.5	0.65 <sup>b</sup>
P-T ICL	0.44	0.38	0.34	3.10
N	16	30	12	

<sup>a</sup>  $F_{.05} (df = 2/54) = 3.17$ .

<sup>b</sup> Four-month scores controlled on this variable since there were no initial scores on this variable.

### Treatment Gains: Eight Months

Table 4 presents initial and 8-month means of the combined groups on the various criteria as well as the results of *t* tests (one-tailed). Of the patient measures, the Ego Strength scale scores increase significantly in the predicted direction. Patients also use significantly fewer Dependency adjectives in describing themselves at the end of 8 months of therapy. There are no significant shifts from pretreatment status on the other patient criteria. As was true at 4 months, the therapists observe a significant decrease in Severity of Illness as compared with the beginning of therapy. In addition, therapists note significantly more IC + SR Changes at 8 months than they had observed at 4 months. Thus both therapists and patients provide evidence of improvement over initial or 4-month status even though this improvement does not relate significantly to frequency of treatment.

### A ONE-YEAR FOLLOW-UP

A high proportion of the original group was known to be in treatment at the end of 12 months, although not all on the originally as-

signed treatment schedule. Yet it seemed useful to retest those remaining in order to assess any changes. Thus 1 year after starting psychotherapy, the 133 patients who had completed at least 4 months of psychotherapy were retested or asked to return for a follow-up re-evaluation. Of 102 patients retested (77% of the total) 55 had remained in psychotherapy for the entire year (Ingroup), and 47 had completed or terminated psychotherapy between 4 months and 1 year (Outgroup). The 31 patients who did not respond to the follow-up request did not differ significantly in background characteristics from the re-evaluated group.

To test for effects of treatment frequency at the one year follow-up, the analysis required a grouping of patients which differed from the original experimental design. By one year, too few patients remained in therapy on the original randomly assigned treatment schedules for meaningful analysis by treatment frequency group. It also seemed essential to analyze the In- and Outgroups separately because of different duration of treatment. Therefore, both the Ins and Outs were dichotomized into two groups on the basis of the number of interviews received. Patients

TABLE 4  
CHANGES OVER AN EIGHT-MONTH PERIOD FOR  
COMBINED GROUPS

Criterion	Means		t <sup>a</sup>
	Initial	8 month	
Patient measure:			
Manifest Anxiety	31.1	30.0	1.17
Ego Strength	31.0	32.4	2.40**
Symptom Checklist	8.0	7.4	0.53
Self-Rating	44.8	45.7	0.86
Sociability	6.4	6.4	0.22
Friendliness	8.7	8.7	0.26
ICL: Assertiveness	9.4	9.3	0.37
ICL: Cooperativeness	8.4	8.0	1.38
ICL: Hostility	12.3	12.0	0.65
ICL: Dependence	16.4	15.0	2.83**
Therapist measure:			
Severity of Illness	21.1	19.8	2.35**
IR	8.6 <sup>b</sup>	8.8	0.28
IC + SR	10.2 <sup>b</sup>	12.1	1.88*
N	58		

<sup>a</sup>  $t_{.05} \text{ (one-tailed)} = 1.64$ .

<sup>b</sup> Four-month means.

\*  $p < .05$ .

\*\*  $p < .01$ .

receiving less than the median number of interviews were classed as Lows; all others were classed as Highs. Comparisons of High and Low interview groups were then made separately for both Ins and Outs. The actual effect of dichotomizing was that all cases assigned to twice weekly treatment fell into the High interview subgroups, all cases originally assigned to biweekly treatment fell into the Low interview subgroups, but the once weekly group patients were distributed to both groups depending on the number of appointments they actually kept.

### *Ingroup Comparison*

The High group consisted of 28 patients seen an average of 62 interviews, with a range of 43-96 sessions. The Low group consisted of 27 patients seen an average of 29 interviews, with a range from 17-42 interviews. The High and Low groups differed significantly on a few variables before starting treatment. The High interview group was younger, included more unemployed patients, more single men, and reported fewer symptoms and complaints at the start of therapy. Therapists also expressed significantly greater interest in the types of problems presented by the High group. Thus it appears that the 12-month comparison groups are not as comparable on initial status as the 4- and 8-month comparison groups which differed very little. This could well be due to the fact that the High and Low groups were formed partly on the basis of the number of appointments the patients actually kept during the year of treatment.

Table 5 presents the results of tests for differences between the High and Low groups at one year. As previously, initial scores are controlled by analysis of covariance. Two new patient criteria—Social Changes and Psychological Changes—are also listed in Table 5. The two measures contain a total of 19 items taken from the therapist IC + SR measure and adapted for patient self-reports. As the names imply, one contains a set of changes in the social-interpersonal adjustment area, and the other contains changes in self-attitudes and psychological symptoms. None of the therapist measures even suggests a difference between Highs and Lows at one year. Only

TABLE 5  
ADJUSTED TREATMENT GROUP MEANS FOR PATIENTS  
IN THERAPY AT TWELVE MONTHS  
(Initial scores controlled)

Criterion	Interview frequency		F <sup>a</sup>
	High	Low	
Patient measure:			
Manifest Anxiety	26.0	26.9	0.22
Ego Strength	34.9	32.8	1.05 <sup>b</sup>
Symptom Checklist	7.3	6.7	0.49
Self-Rating	48.4	45.9	1.37
Sociability	6.9	6.9	0.00
Friendliness	8.4	9.9	5.23 <sup>*</sup>
ICL: Assertiveness	10.1	9.7	0.17
ICL: Cooperativeness	7.2	8.8	8.62 <sup>**</sup>
ICL: Hostility	12.0	10.8	1.34
ICL: Dependence	14.9	13.6	1.35
Social Changes	4.4	3.4	1.69 <sup>c</sup>
Psychological Changes	4.9	4.7	0.06 <sup>c</sup>
Therapist measure:			
Severity of Illness	18.8	18.9	0.03
IR	8.0	9.0	0.00 <sup>d</sup>
IC + SR	16.3	16.4	0.00 <sup>e</sup>
P-T ICL	0.36	0.36	0.01
N	28	27	

<sup>a</sup>  $F_{.05} (df = 1/52) = 4.04$ .

<sup>b</sup> Twelve-month means presented and tested, as covariance analysis assumptions untenable.

<sup>c</sup> Twelve-month means presented and tested as there were no initial scores on this variable.

<sup>d</sup> Median changes listed for each group.  $\chi^2$  for median test is presented as variances were heterogeneous.

<sup>e</sup> Four-month scores controlled as there were no initial scores.

<sup>\*</sup>  $p < .05$ .

<sup>\*\*</sup>  $p < .01$ .

two patient measures indicate a significant effect of number of interviews. The High patients scored significantly lower on the Friendliness scale and used significantly fewer Cooperativeness-Responsibility Adjectives in describing themselves.

The labels Friendliness and Cooperativeness probably do not convey very well the nature of the changes that took place. An item analysis revealed that the High group showed reduced Friendliness because they more often rejected such platitudes as "It pays to turn the other cheek rather than to fight," and more often endorsed such statements as "It is often necessary to fight for what is right," and "If I resent someone's actions I promptly tell him so." The reduced Cooperativeness scores were due to less frequent endorsement of such self-descriptions as "Accepts advice readily," "kind and reassuring," and "often helped by others." Item analysis, thus, suggests that the High group



did not become unfriendly and uncooperative as the test labels would suggest, but instead became more outspoken, assertive, independent, and determined to protect their own interests.

### Outgroup Comparison

The High interview group consisted of 23 patients no longer receiving treatment, seen an average of 34 interviews with a range of 23–63 sessions. The Low group consisted of 24 patients seen an average of 16 sessions, with a range from 10–22 interviews. There was no difference in duration of treatment for the High and Low groups; mean length of therapy was 29 weeks for both groups. Tests to determine if the High and Low groups were comparable samples of patients when treatment began revealed only one significant difference on background characteristics and initial criterion scores. Therapists judged the High group to be significantly more severely ill than the Low group. (Means: 22.5 vs. 18.8,  $t = 2.71$ ,  $p < .01$ .) Analysis of the seven items comprising the set of severity scales indicated therapists judged the High group to be more distressed, more suspicious of others, and more self-preoccupied. They rated the High group as moderately ill and the Low group as mildly-to-moderately ill.

TABLE 6

ADJUSTED TREATMENT GROUP MEANS FOR PATIENTS  
OUT OF THERAPY AT TWELVE MONTHS  
(Initial scores controlled)

Criterion	Interview frequency		$F_a$
	High	Low	
Manifest Anxiety	24.9	28.4	2.86
Ego Strength	33.8	33.2	0.14
Symptom Checklist	7.7	7.5	0.02
Self-Rating	49.2	47.7	0.59
Sociability	8.0	9.0	2.42
Friendliness	7.9	10.0	6.60*
ICL: Assertiveness	10.3	11.3	1.49
ICL: Cooperativeness	7.9	9.2	5.02*
ICL: Hostility	11.3	10.9	0.07
ICL: Dependency	11.6	14.7	6.94**
Social Changes	2.9	3.0	0.01 <sup>b</sup>
Psychological Changes	2.8	4.1	3.29 <sup>b</sup>
<i>N</i>	23	24	

<sup>a</sup>  $F_{.05} (df = 1/44) = 4.07$ .

<sup>b</sup> Actual twelve-month means are presented and tested as there were no initial scores on this variable.

\*  $p < .05$ .

\*\*  $p < .01$ .

TABLE 7

INITIAL AND TWELVE MONTH TEST MEANS FOR  
PATIENTS IN THERAPY AT TWELVE MONTHS

Criterion	Means		$F^a$
	Initial	12 month	
Patient measure:			
Manifest Anxiety	29.9	26.4	6.61 <sup>b**</sup>
Ego Strength	31.3	33.8	4.26 <sup>**</sup>
Symptom Checklist	8.0	7.0	1.94*
Self-Rating	45.7	47.1	1.29
Sociability	6.9	6.9	0.11
ICL: Assertiveness	9.5	9.9	0.75
ICL: Hostility	11.8	11.4	0.68
ICL: Dependency	15.1	14.3	1.45
Therapist measure:			
Severity of Illness	21.5	18.9	4.86 <sup>**</sup>
IR	8.6 <sup>c</sup>	8.9	0.52
IC + SR	11.9 <sup>c</sup>	16.4	3.20 <sup>b*</sup>
P-T ICL	0.35	0.36	0.35
<i>N</i>	55		

<sup>a</sup>  $t_{.05} (one-tailed) = 1.64$ .

<sup>b</sup> Table entry is  $\chi^2$  for McNemar (1955) test of changes.  $t$  test invalid because of heterogeneous variances.  $\chi^2_{.05} (one-tailed) = 2.71$ .

<sup>c</sup> Four-month means.

\*  $p < .05$ .

\*\*  $p < .01$ .

Table 6 presents the results of tests for differences between the High and Low Outgroups at one year. No therapist measures were available as the Outs were no longer receiving therapy. Three of the differences on the patient measures are significant at  $p < .05$ . Two of the differences—Friendliness and Cooperativeness—involve the same variables in the same direction as found in the Ingroup analysis. In addition the High group patients showed a significantly greater reduction in number of Dependency Adjectives used in self-descriptions. The Low patients actually showed a slight nonsignificant increase in the number of such adjectives. Examples of self-descriptions used less frequently by the High interview group are obeys too willingly, shy, wants everyone's love, and likes everybody. All three significant differences tend to confirm the findings with the Ingroup, that patients with more interviews see themselves as more assertive and outspoken in pursuing their own interests, less dependent, and less willing to be imposed upon by others. In addition to the above differences, the greater reduction in MA approaches significance ( $p < .10$ ) and lends support to the impression of greater improvement for the High inter-

view patients. However, no simple picture of improvement can be claimed for the High patients. Both groups report about the same number of symptoms and complaints (Symptom Checklist) at 12 months, and the Low group tends to report slightly more Psychological Changes ( $p < .10$ ) since they started therapy.

### *Ingroup Treatment Gains*

Table 7 presents tests of differences between initial and 12-month means on those patient and therapist measures which were unaffected by interview frequency. Significant changes from pretreatment status occurred on three of the patient measures. *MA* scores and the number of reported symptoms decreased, while the Ego Strength scale score increased. Except for the unchanged Sociability scale, the remaining patient measures show slight movements in the predicted direction. Therapists corroborate the pattern of improvement by reporting a decrease in severity of illness that is significant. They also report significantly more interpersonal changes and symptom reductions compared with changes during the first 4 months of psychotherapy. No additional change in the interview relationship is noted after the first 4 months which suggests that this measure may be sensitive only to early changes. In summary, the patients in treatment at the close of 1 year show a pattern of favorable change from initial status that is somewhat broader than the pattern at 8 months and considerably broader and more consistent than at 4 months.

### *Outgroup Treatment Gains*

Tests were made to determine whether there were significant shifts from pretreatment status for the entire Outgroup on the variables which did not show a significant relation to treatment frequency. Comparisons of initial and 12-month means were made on the following variables: Ego Strength, Symptom Checklist, Self-Rating, Sociability, Assertiveness, and Hostility. None of the differences was significant. Thus, among the Outgroup, only those patients with a high number of interviews present any evidence of change from pretreatment status.

## DISCUSSION

### *Related Studies*

Imber et al. (1957) have reported a study similar to the present investigation in objective. They hypothesized that patients having fewer and briefer sessions of psychotherapy would show significantly less improvement than patients with more and longer sessions over the same period of time. Patients were assigned at random to group psychotherapy for 1.5 hours once a week, to individual psychotherapy for 1 hour once a week, and to minimal contact therapy for .5 hour once every two weeks. After 6 months they report more improvement in group and individual psychotherapy patients than in minimal contact patients. Unfortunately length of interview, frequency of treatment and type of treatment are confounded in the study design. It is not possible to infer which variable was the effective agent.

Ends and Page (1959) claim to have demonstrated that doubling the number of group therapy sessions with a constant time interval increases therapeutic movement. Lack of statistical or experimental control of sampling sources and other defects cast serious doubt on their findings. Cartwright (1955) and Taylor (1956) have independently found that ratings of success or improvement on closed cases increases with number of treatments received. Duration of treatment seems not to have been controlled in either study. Further these studies are concerned with closed cases rather than open cases as in the present report.

### *Contingent Variables*

Ordinarily therapist and patient mutually decide how frequently treatment sessions should be scheduled. Yet random assignment of patients to schedules without regard to judged suitability may have effected outcome. The hypothesis that patients on schedules rated suitable by the therapist after the initial interview would improve more than patients on schedules rated unsuitable was tested on 10 study criteria obtained at the 4 month re-evaluation (McNair & Lorr, 1960a). The evidence did not support the hypotheses for any of the criteria. Suitability ratings were, how-



ever, associated with therapists' personal reactions to patients.

The factor of chronicity of illness of the group studied calls for some consideration. The patients studied were all male veterans who were treated for service associated psychiatric problems that date, presumably, from World War II or the Korean War. Illness thus dates back 8-15 years and may be regarded as chronic. Further about half of the study group had been in psychotherapy once before, for a median period of 10 months. Yet a survey by Feldman et al. (1958) showed that the typical Veterans Administration open case is seen for 1.5 years. Hollingshead and Redlich (1958) found the median number of years neurotic patients had been seen in treatment was about a year for Social Class III and Class IV persons. It is thus not surprising that substantial changes did not appear until after 8 months of treatment nor is it unimpressive in view of the chronicity of illness.

#### *Patient Treatment Gains*

In the absence of a control group of patients who receive no treatment, what can be inferred concerning the changes over time reported by patients and therapists? There is scant direct experimental justification for considering the changes obtained as effects of psychotherapy. However, there is much indirect support for the hypothesis that the changes obtained are due to individual psychotherapy. Changes obtained are in predicted directions and are both plausible and consistent. Furthermore, the pattern of change becomes broader at each assessment period. At the end of 4 months few significant changes were obtained. By the close of 8 months there were significant increases in Ego Strength scores and in ICL Dependency scores. Therapists observed a reduction in severity of illness and a reduction in symptoms. After one year in treatment patients exhibited significant gain in Ego Strength and reductions in Manifest Anxiety and Symptom distress. On this occasion the therapists again noted a reduction in severity of illness. However, most of the IC and SR gains were in the area of Interpersonal Relations. Little further reduction in the number and severity of symptoms

was reported. The order of these changes is what one might expect. Symptoms are first reduced and the more complex gains in interpersonal relations come later.

It would be difficult to explain these systematic gains in terms of a hypothesis of spontaneous improvement. For example, the Low Outgroup failed to change although in treatment as long as the High Outgroup. If patients improved spontaneously the Low Outgroup could also be expected to gain unless the number of treatments was an agent. Another likely argument is that patients and therapists have a need to justify the expenditure of time and effort in treatment and respond by reporting favorable changes. This argument is persuasive but it fails to account for the relative absence of changes after 4 months of treatment and the limited changes obtained after 8 months. Why do patients and therapists restrict the changes to certain areas? A tentative inference can thus be made, from the internal evidence, that the gains reported are ascribable to psychotherapy in a mental hygiene clinic setting.

Apparently the nature of the measures also played a role in controlling patient and therapist bias. Therapists were asked to report on specific attitudinal and behavioral changes observed. If they had been asked to indicate the degree of global improvement observed the findings would have been far more impressive and probably much more biased. Likewise, patient gains were determined from status scores. If given the opportunity to report how much they "improved" the changes would have been much larger and also more easily distorted.

#### *Treatment Frequency and Duration*

What may be concluded concerning the effects of increasing the number of interviews per month? There is little evidence that treatment frequency effects treatment outcome for the range of schedules used, for the kinds of patients studied, and for the methods of treatment applied. The results might have been different if, for example, the method of treatment was psychoanalysis and the treatment schedule was three to five times a week. However, this is a conjecture that has no experimental data for support.



The major finding is that for the conditions of this study duration of therapy, rather than treatment frequency, is associated with predicted change over an 8-month period. For the one-year period the basis of treatment effects is uncertain because the analysis is no longer based on randomly assigned experimentally controlled groups. The composition of the High and Low interview groups at 12 months is determined, in part, by the number of appointments patients kept. On the other hand the agreement between the High In-group and High Outgroup in the kinds of changes displayed suggests that the number of treatments may have some influence on outcome. Of the two variables, duration of psychotherapy seems more influential than treatment frequency. The pattern of change becomes broader and more consistent at each assessment period. However, it should be noted that for the 8-month and 1-year periods data was not available from sources outside of therapy such as independent interviewer, spouse, friends, or relatives. Conclusions reached here, therefore, should be qualified by this limitation on sources of evidence.

Only the data for patients out of therapy before the end of a year support Cartwright's (1955) finding that the number of treatments has a nonlinear relation to therapeutic gain as reported by therapists. The data on patients out of treatment suggest that a minimum number of treatments are needed to effect a determinable change. The Low Outgroup did not change over the one-year period while the High Outgroup did. Mere passage of time is thus not sufficient to account for this difference.

The evidence suggests that change requires time. Perhaps trial and error testing is a prerequisite for the process of growth and change. New ways of reacting interpersonally must be tested repeatedly in natural settings before what has been learned becomes consolidated. Insights must be put into practice. The findings of this study suggest that traditional treatment frequency schedules be examined. If duration is the more influential variable for ordinary psychoanalytic type psychotherapy, Therapist time can be spread over more patients with fewer contacts and at less cost to patients.

## SUMMARY

1. The hypothesis that patient changes in predicted directions will increase with the number of treatment interviews is not supported for the first 4-month period of treatment. Patients, therapists, and social work interviewers agree that no differences are ascribable to treatment frequency.

2. A comparison of treatment frequency groups after 8 months of treatment on patient and therapist measures again fail to support the hypothesis that the number of treatments is positively related to therapeutic gain.

3. After 4 months of treatment both therapists and social workers observe an overall decrease on severity of illness not associated with frequency of interviews. Social workers also note greater patient Social Adjustment.

4. Following 8 months of psychotherapy patients increase significantly in Ego Strength score, as predicted. Patients also describe themselves with fewer Dependency Adjectives than at pretreatment. Neither change is associated with frequency of therapy interviews.

5. Following 8 months of psychotherapy there is a significant decrease in severity of illness not associated with treatment frequency. Therapists also report significantly more interpersonal changes and reductions in symptoms than for the 4-month period.

6. The High interview group patients who remain in psychotherapy for a year describe themselves as more outspoken, assertive, independent, and determined to protect their own interests than the Low group patients.

7. The High interview group patients out of treatment at the one-year follow-up describe themselves as more assertive, outspoken, and independent than the Low group. In addition, the High group shows a greater reduction in the use of Dependency Adjectives.

8. As a group those patients who remain in psychotherapy for a year change on certain test variables unaffected by treatment frequency. They decrease significantly in anxiety (*MA* scale), report fewer symptoms, and increase in ego strength (*Barron* scale). Therapists report significant decreases in severity of illness, and a greater number of favorable interpersonal changes and symptom reductions as compared to the first 4 months of treatment.

ment. The pattern of favorable change is broader than at 4 and 8 months.

9. Of patients out of treatment at the time of the one-year follow-up re-evaluations, only the High interview group exhibits changes from pretreatment status.

#### REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSOHN, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- APPLEZWEIG, M. H., DIBNER, A. S., & OSBOURNE, R. L. PEAQ: A measure of psychopathic behavior. *J. clin. Psychol.*, 1958, 14, 26-30.
- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.
- CARTWRIGHT, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, 19, 357-363.
- ENDS, E. J., & PAGE, C. W. Group psychotherapy and concomitant psychological change. *Psychol. Monogr.*, 1959, 73(10, Whole No. 480).
- FELDMAN, R., LORR, M., & RUSSELL, S. B. A mental hygiene clinic case survey. *J. clin. Psychol.*, 1958, 14, 245-250.
- GUILFORD, J. P., & ZIMMERMAN, W. *The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretation*. Beverly Hills, Calif.: Sheridan Supply, 1949.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness: A community study*. New York: Wiley, 1958.
- IMBER, S. D., FRANK, J. D., NASH, E. H., STONE, A. R., & GLIEDMAN, L. H. Improvement and amount of therapeutic contact: An alternative to the use of no treatment controls in psychotherapy. *J. consult. Psychol.*, 1957, 21, 309-315.
- LA FORCE, R., & SUCZEK, R. The interpersonal dimension of personality: III. An interpersonal checklist. *J. Pers.*, 1955, 24, 94-112.
- LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.
- LORR, M., HOLSOFFLE, J. D., & TURK, ELIZABETH. A measure of severity of illness. *J. clin. Psychol.*, 1956, 12, 384-386.
- MCNAIR, D. M., & LORR, M. Therapist judgments of appropriateness of psychotherapy frequency schedules. *J. consult. Psychol.*, 1960, 24, 500-506. (a)
- MCNAIR, D. M., & LORR, M. Two therapist measures of patient change in psychotherapy. *Amer. Psychologist*, 1960, 15, 386. (Abstract) (b)
- MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.
- MYERS, J. K., & AULD, F. Some variables related to outcome of psychotherapy. *J. clin. Psychol.*, 1955, 11, 51-54.
- SEEMAN, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 99-108.
- TAYLOR, J. W. Relationship of success and length in psychotherapy. *J. consult. Psychol.*, 1956, 20, 332.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- THOMPSON, CLARA. *Psychoanalysis: Evaluation and development*. New York: Hermitage House, 1950.

(Received March 2, 1961)

# THE STIMULUS QUALITIES OF THE SCAPEGOAT<sup>1</sup>

LEONARD BERKOWITZ AND JAMES A. GREEN

*University of Wisconsin*

Generally speaking, most explanations of social prejudice are somewhat one-sided. They concentrate on factors either in the prejudiced individual or in the victimized group, but typically do not effectively relate the attacker to the attacked. The present paper will attempt to show that the object serving as the target for the intolerant person's aggression usually has certain stimulus qualities for this person, and that objects not possessing these characteristics are less likely to be attacked. In dealing with ethnic prejudice, in other words, it is necessary to consider both the aggressor and the available targets. To focus on either alone is to give only part of the picture.

Such one-sidedness is particularly apparent in the scapegoat theory of prejudice. The details often vary from one writer to another, but all versions of this common social science doctrine seem agreed on at least the following features: Frustration generates aggressive tendencies, which cannot be directed against the actual thwarting agent because this agent is not visible, or is capable of retaliating with severely punitive actions. A needed outlet is then found for the pent-up aggressive "energy" through attacks upon some innocent minority group. The displaced aggression is rationalized by blaming the minority for the frustrations the aggressor has experienced, and/or attributing undesirable characteristics to this group.

Several authorities (e.g., Allport, 1954; Zawadski, 1948) have noted that the above type of theorizing leaves many important questions unanswered. It does not tell us, for example, why a particular minority is attacked when any number of groups are available. Why are Jews more likely to be the victim of the displaced aggression than, say, people of Scottish descent? Considerations

other than that an individual is frustrated and unable to attack the source of his frustration obviously must be introduced in order to handle this problem. As Zawadski (1948) pointed out, analyses of the scapegoating process tend to be "pure drive" theories, explaining the origin of the aggressive tendencies, but not the target selection. Other variables must be employed to deal with the choice of object for aggression.

Many writers (e.g., Williams, 1947), basing their reasoning on the psychoanalytic "energy" model of behavior, have assumed, almost as a matter of course, that the thwarted individual who is afraid to attack the actual anger instigator will aggress against the person least likely to harm him by retaliatory aggression. An aggressive outlet must be found and presumably is found in attacks upon objects who cannot inflict injury in return. The most likely target for displaced aggression, then, supposedly is the safest available target. This formulation has also been seriously questioned (e.g., Allport, 1954). White and Lippitt (1960) have been among the most recent critics of this "safety" hypothesis in their latest report of their now classic leadership study. They observed a number of instances of scapegoating in the frustrated autocratically led groups, but claimed that the victims were never the weakest or most passive boys in the club. The boys singled out for aggression in one autocratic group "were both boys who could hold their own against any of the others taken singly," while in another club the scapegoat was the largest and heaviest boy. White and Lippitt suggested that these particular boys were attacked in an attempt to recover status or self-esteem. Frustrations, they maintained, elicit hostility only when they lower self-esteem. By aggressing against these boys who were fairly strong and dangerous but without being excessively formidable, the attackers presumably could regard themselves as strong and potent in their own right, and thus their

<sup>1</sup> This research was supported by Research Grant M1540(C3) from the National Institute of Mental Health, United States Public Health Service.



self-esteem supposedly would be restored (p. 166).

It is doubtful, however, whether status recovery can satisfactorily account for every case of target selection in displaced aggression. Why would attacks upon Jews provide a greater restoration of self-esteem than attacks upon minorities of Nordic origin? The solution to this problem of object choice must involve the stimulus properties of the various available objects. Along these lines, Williams (1947), among others, proposed that scapegoats frequently are visibly different or strange. It is the strangeness of the available objects that determines their likelihood of evoking displaced hostility. Strangeness or difference itself supposedly is disturbing. Allport (1954) has been inclined to accept such a thesis, contending that whatever instinctive basis there may be for group prejudice can perhaps be found in the "hesitant response . . . human beings have to strangeness" (p. 130). Babies of about 6 months of age cry or become emotionally upset when a stranger draws near them, and such a reaction tendency may persist into adulthood. The present writers, believe, nevertheless, that strangeness is upsetting only under certain limiting conditions. Animal research (cf. White, 1959) and the rush of tourists to foreign countries indicate that strange and novel stimuli may be enticing in some circumstances.

Fear of a stranger largely arises when the individual expects the unknown person to be potentially dangerous. If a person is afraid of strangers or people who are greatly different from himself he probably views most people as being dangerous; the stranger is an "ink-blot" eliciting the responses the person customarily makes to people. Ethnocentric personalities, of course, are relatively likely to be antagonistic to those who are different (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), and evidence suggests these people often are fairly insecure. They supposedly are uncertain of themselves, the world about them, and their place in the world. Thus, according to a study by Allport and Kramer (1946), highly prejudiced adults are much more likely than their more tolerant peers to agree that, "The world is a hazardous place in which men are basically evil and danger-

ous." Having this outlook, it is not surprising that they are unfriendly toward outgroups. Since the world is a threatening place for them, an unknown person from this world also is potentially dangerous.

The above analysis explains the ethnocentric individual's relatively strong tendency to generalize his frustration induced aggressive tendencies toward strangers (Berkowitz, 1959). The stranger, and more generally, the alien group, is somewhat threatening and, because it is threatening, is disliked. Dislike, we contend, mediates the generalization of aggression. *Aggression will generalize from the anger-instigator to another person in direct proportion to the degree of dislike for this latter individual.* If this hypothesis is correct, we would have a means of integrating the scapegoat theory of prejudice with those other accounts of intergroup conflict focusing upon the characteristics of the socially victimized ethnic groups. Where the scapegoat theory is a "pure drive" theory, to employ Zawadski's (1948) terminology, these latter notions—such as the so-called "well-earned reputation" doctrine—can be described as "pure stimulus" theories (Zawadski, 1948). Though they differ in important ways, these "stimulus" theories have at least one aspect in common; they all provide reasons why given minority groups are disliked. Putting it simply, the present argument holds that the aggressive tendencies engendered by frustrations are generalized to those groups whose perceived characteristics result in their being disliked. The individual may absorb his family's or his culture's negative attitudes towards a particular group, or he may have had unpleasant experiences with members of this group. The genesis of the negative attitude is unimportant as far as we can see. As long as a group is disliked, whatever the reason, it is a likely target for displaced aggression.

This is not to say that particular characteristics of the minority group have no part in determining its likelihood of becoming a scapegoat. While investigators of the authoritarian personality usually maintain only that a group is attacked merely because it is an "outgroup" (cf. Adorno et al., 1950, p. 233), they sometimes, in company with other psychoanalytically oriented writers, also empha-

size the importance of the group's perceived qualities. The prejudiced individual, for example, supposedly projects his own disapproved sexual wishes onto Negroes because the stereotype of this group easily accommodates such a projection, and then hates Negroes because of their perceived sexuality. Similarly, Jews are said to symbolize other properties the authoritarian personality unconsciously sees and detests in himself.

The problem we are addressing ourselves to is the theoretical significance of the characteristics attributed to (or actually possessed by) the outgroups. Jews share few, if any, specific features with Negroes. In the United States at least, sexual qualities typically are not assigned to the former (Allport, 1954), but both groups are likely targets for the same prejudiced individual's hostility. Just what do Jews, Negroes, and other outgroups have in common that results in their all being victimized? Most investigators of authoritarianism have been too concerned with the detailed depths of the prejudiced personality to look for the abstract principle that effectively integrates the various instances of scapegoating.

The specific characteristics perceived in a group do four things from our point of view. Most important, they determine whether the minority is disliked and, if so, how strongly. This is the quality shared by the victims of displaced hostility. They are disliked for different reasons (the *Authoritarian Personality* and other writings give us some of these reasons), but all are detested. As a result of being the object of negative attitudes, then, hostility engendered by some frustration will generalize fairly readily to these outgroups. Two, the extent of hostility generalization, we believe, is a function of the total degree of association between the immediate frustrator and the objects available for scapegoating. The perceived properties of these latter objects, as well as the dislike for them, contribute to the psychological ties the thwarted person can draw between his frustrator and the available targets. The generalized aggressive tendencies of course will not lead to overt attacks if the intolerant person fears he will be punished for displaying aggression and/or believes such hostility is morally improper.

The outgroup characteristics also may affect these factors. They can determine, three, whether the prejudiced individual believes it is safe to attack a given outgroup and, four, whether he is ethically justified in doing this.

The present reasoning obviously is based on the stimulus generalization of displacement (cf. Miller, 1948). Although our specific predictions do not necessarily require postulating stimulus generalization, we assume the negative attitudes associated with both the immediate frustrator and the disliked group give rise to a generalization continuum linking these two objects. In other words, because the thwarted individual makes the same implicit responses to the two objects the disliked group is associated with the frustrator. There may be an association between the two solely because they have elicited the same negative emotions in the individual. Another possibility is that the frustrated person implicitly applies the same label to both disliked objects, placing them in the same negatively evaluated category. Dollard and Miller (1950), in advancing the somewhat similar concept, "acquired equivalence of cues," contended that a previously neutral stimulus will produce the response elicited by a particular category of stimuli after the subject has learned to apply the category name to this stimulus. But however it comes about, there presumably is an acquired (i.e., response induced) equivalence between the frustrator and the minority group which mediates the generalization of aggressive responses from the former to the latter.

Two earlier papers by Berkowitz and Holmes (1959, 1960) provide evidence supporting the "dislike hypothesis." In the first, findings were reported suggesting that hostility is indeed generalized from the immediate frustrator to another stimulus person (P) the subject had previously learned to dislike. Subjects were first induced either to like or dislike their experimental partners (the Ps). After this, half of the subjects in each of these conditions were frustrated by the experimenter, with the others receiving a more pleasant treatment from him. Then, in the last phase of the study, pairs of subjects (each subject regarding the other as P) were brought together again for a cooperative task.



It was shown that the subjects who had expressed relatively intense hostility to the experimenter after having been frustrated by him and who then were assembled with a disliked P on the final task generally expressed the strongest unfriendliness to this P on a questionnaire evaluation of him at the end of the session. The presumably intense hostility engendered by the experimenter apparently had generalized to the disliked P to a greater extent than to the more highly liked partner.

The second experiment (Berkowitz & Holmes, 1960) obtained similar findings with stronger and more direct acts of aggression. Pairs of subjects were put through the same procedures employed in the first study, but this time during the last phase of the experiment each subject was given a socially sanctioned opportunity to administer electric shocks to his partner. The subject was to evaluate P's performance on an assigned task by giving P electric shocks: one if the product was very good, more than this if the performance was thought to be poor. In actuality, each subject was shown the same product. There was the greatest increase in the number of shocks administered to the partner (in comparison to the number given during a baseline period at the start of the study) when the subject had been frustrated by the experimenter and then had an opportunity to shock the disliked P. Again it seems as if the aggressive tendencies evoked by the unpleasant experimenter has transferred to the unpleasant P.

However, there is at least one important difficulty confronting this interpretation. The subjects sending the greatest number of shocks to P had been frustrated twice: once during the manipulation creating the dislike to P, and again by the experimenter. Although internal evidence contrary to this possibility was reported, it is conceivable that the relatively great amount of aggression exhibited by these subjects was due solely to the intense anger aroused within them by the two thwartings. They could have been so angry they would have attacked anyone strongly, whether this target was disliked or not.

The present experiment seeks to test this alternative explanation. Essentially the same procedures utilized in the earlier experiments

are again employed. However, the subject is now given two people to evaluate during the final phase of the experiment. One of these, as in the first two studies, had previously been either friendly or unfriendly to him. The second person is presumably fairly neutral since the subject had not interacted with him before. If the intense anger created by the two thwartings was the crucial determinant of the previous results, there should be no difference in the final evaluation of these two stimulus people, and both should be regarded more unfavorably after the subject is frustrated twice than after the subject is thwarted once or not at all. On the other hand, only the disliked P should receive the harshest evaluation, and not the "neutral" stimulus person, after the subject is thwarted by the experimenter if the existing attitude towards the available target affects this object's likelihood of receiving generalized aggression.

## METHOD

### *Subjects*

The subjects were male students from introductory psychology classes at the University of Wisconsin who volunteered without knowing the nature of the experiment. Four subjects were discarded after pretesting was terminated—one because he had not completed the final questionnaire and three who indicated they were suspicious of the treatment accorded them. No more than two of these "discards" came from any one condition. There were 18 subjects in each of the four experimental conditions in the final sample.

### *Procedure*

Two subjects who did not know each other were scheduled for any one experimental period. After both had arrived at the laboratory they were joined by a third male posing as the third experimental subject but who was, in actuality, the experimenter's confederate. The experimenter explained the ostensible purpose of the study, saying the experiment was to investigate the effects of stress upon creativity. They were told each subject would first make a judgment of the personalities of his two partners (supposedly because creative people were good at making "snap judgments"), and then two of the three would work on a problem solving task under mild stress, while the third person would be alone in a natural condition as a "control." The stress, they were told, would come from knowing they might receive several electric shocks if their performance was not too good. Each of the two people in the Stress condition was to work independently of the other on the assigned task. When they had



finished they were to exchange their problem solutions so that each would serve as the judge of the other's performance. Each subject would transmit his evaluation of his partner's product by giving him electric shocks. There would be one shock if the product was very good and more than this if it was thought to be inadequate. Following this, the experimenter said, each subject would work alone and he would be judged (without shocks) by the experimenter. Finally, in the last part of the session, the three men would work together on a group task.

At this time the subjects were told the shocks would be relatively mild and they were given an opportunity to withdraw from the experiment if they objected to receiving shocks. None did. Letters were then assigned to the three men, the two "real subjects" being called A and C, and the confederate B. A and C were told that they would work under the Stress condition, and all three men were led to separate rooms where the naive subjects indicated their first impressions of their partners on an adjective checklist. The code letters were used in all of the ratings made throughout the experiment.

When the first personality evaluations were completed, both men were given the task of designing a "novel, imaginative, and creative" floor plan for a house. Each subject was informed that he would have 5 minutes to work on this problem and that his partner was to have the same task. After 5 minutes had passed by, the experimenter collected the subject's work, strapped shock electrodes onto his wrist, and then showed him what was supposedly the partner's (P's) performance but actually was previously constructed to be standard for all conditions. Each subject was told he was to go first in evaluating the other's performance by means of the electric shocks. He was to press a button on a nearby table one or more times depending upon his judgment of the other's work. The experimenter then left the room ostensibly to deliver the subject's product to his partner. When his instruments (to which the shock buttons were connected) informed him that both subjects had given shocks, the experimenter administered the shocks the subjects believed came from their partners.

### *Experimental Manipulations*

The first manipulation, as in the preceding investigations, was designed to create differences in initial liking for P. Subjects in the Initial Liking for P condition received one shock, indicating that P had given them the most favorable evaluation. The other half of the subjects, those in the Initial Dislike for P condition, were given six electric shocks. Thus, on top of whatever physical hurt they felt, these subjects knew P had derogated their performance.

Subjects worked alone on the next problem (suggesting an original idea for attracting new customers to a gasoline station) without exchanging papers, but this time the experimenter, rather than P, was the perceived anger instigator. Half of the subjects

in each of the above two conditions (those receiving the Frustrated by the Experimenter "treatment") were criticized and insulted by the experimenter during this problem by standardized messages transmitted to them via earphones. The remaining subjects (in the Nonfrustrated by the Experimenter condition) heard a friendlier evaluation of their work. Following this interaction with the experimenter, subjects were asked to fill out a short questionnaire, supposedly an evaluation of psychological experiments to go to the Chairman of the Department. This form, of course, was primarily intended to test the success of the manipulation (Frustration by the Experimenter) in arousing unfriendliness toward the experimenter.

In the third and final part of the study the two subjects and the confederate were brought together for 5 minutes to assemble a footbridge from materials stacked in the room. The three men returned to their individual rooms at the end of the work period. Once there, the subjects completed an alternate form of adjective checklist indicating their assessments of their two peers supposedly based on all the information they had obtained throughout the experiment.<sup>2</sup> When this was done, the experimenter explained the actual purpose of the experiment and informed them of the deceptions he had practiced. Many of the subjects expressed a good deal of interest in the study and all promised not to talk about it to their friends.

### *Dependent Variables*

The measure of each subject's attitudes toward P (with whom he had interacted throughout the study), and toward the confederate (the presumably neutral person) was based on the adjective checklist, a version of a technique used with apparent success in other research in the present program (e.g., Berkowitz, 1960). As mentioned above, two alternate forms were employed, one at the beginning, the other at the end of the session. Each form consisted of 33 adjectives. The subjects in responding to these forms were to mark a True-False IBM sheet

<sup>2</sup>The subjects also filled out a brief four-item scale assessing their attitudes toward each of their two peers. The results with this measure, not reported here because of indications that the adjective checklist had affected these later scale responses in the most strongly aroused condition (cf. Berkowitz & Holmes, 1960), are generally consistent with the adjective checklist findings. Thus the Initial Dislike for P-Frustrated by the Experimenter group was the only group experiencing some thwarting which gave P significantly more unfavorable ratings than as assigned to him in the Initial Liking for P-Not Frustrated by the Experimenter condition. The harsh treatment given the subjects by P, furthermore, did not significantly affect the attractiveness of the confederate on this measure, although (as was also found with the adjective checklist scores) there was some hostility generalized to the confederate when the experimenter had been a frustrator.

(using separate IBM sheets but the same form for P and the confederate), stating which of the adjectives characterized the given stimulus person and which did not. Unknown to the subjects, a large group of judges had previously scaled the adjectives as to their overall social desirability. Thus it was possible to assess the level of unfriendliness in the subject's judgments of each of the other two men by adding the number of undesirable traits attributed to the person to the number of favorable characteristics he was said not to possess. As a working assumption, high unfriendliness is taken to signify relatively intense aggressive tendencies.

Two questions embedded in a group of four were employed in assessing the success of the Frustration by the Experimenter. One asked, "How much did you enjoy the experiment?" and the other, "How favorable was your reaction to the experimenter?" In answering each question the subjects were to place a mark at an appropriate position on a linear rating scale anchored at each end by a suitable phrase (e.g., "not at all"). The scores were distances from the favorable end of the continuum in 1 centimeter units. Subjects criticized and insulted by the experimenter should enjoy the experiment less and have a more unfavorable opinion of the experimenter than subjects receiving a friendlier treatment from him.

## RESULTS

### *Success of the Experimental Manipulation*

There were no measures taken in the present study of the subjects' feelings toward P immediately after they had obtained his first "evaluation" of their work. Results from the preceding investigations in this series (Berkowitz & Holmes, 1959, 1960), which employed virtually the same procedure, suggest, nevertheless, that the shock "evaluations" probably affected the level of the subjects' unfriendliness toward P. The subjects in these earlier experiments responded to a questionnaire within a few minutes of receiving the first shocks from P. Those getting the most shocks were more likely than the subjects getting one shock to indicate the P was "unfair," and to say that they had relatively little desire to know this person better. However, since these first investigations employed a greater number of shocks in the Initial Dislike for P condition, the earlier subjects receiving this treatment may have been more strongly angered than the comparable subjects in the present study.

The questionnaire ratings of the experimenter obtained in this experiment clearly in-

dicate that his criticisms and insults in the Frustrated by the Experimenter condition had succeeded in provoking the subjects. Analyses of variance on each of the two relevant measures revealed significant main effects ( $p < .01$  in both cases) for the frustration manipulation and no significant interactions. Thus the "aroused" subjects expressed reliably less enjoyment of the experiment and a significantly more unfavorable evaluation of the experimenter than the subjects receiving friendlier treatment from him.

### *Level of Unfriendliness toward P and the Confederate*

The results obtained with the alternate forms of the adjective checklist administered at the beginning and end of the experimental session are summarized in Table 1.

It can be seen that the two stimulus people were not equally attractive to the subjects at the start of the study; the subjects in each condition had a significantly more favorable "first impression" of P than of the confederate. However, these initial attitudes apparently were only tentative and not too strongly held. Most subjects exhibited a considerable decline in unfriendliness to both of their partners by the end of the experiment.<sup>3</sup>

This decreased unfriendliness, nevertheless, did not occur to the same extent in all conditions. As we would expect from the earlier investigations, the only stimulus people not getting reliably more favorable evaluations at the end of the session were those men whom the subjects presumably had previously learned to dislike and who were being judged by subjects frustrated by the experimenter. Both of these independent variables had to combine in order to retard the growth of friendship. The subjects receiving only one of the harsh treatments, whether from P or the experimenter, showed a considerable decline in unfriendliness toward P. The table also indicates this impeded friendship development was not due simply to an accumulation of frustration effects. Subjects insulted

<sup>3</sup> The decreased unfriendliness may have been merely an effect of differences in adjective checklist forms. We think this is unlikely since the two forms were quite comparable in terms of the overall favorability of the traits listed.



TABLE 1

MEAN UNFRIENDLINESS TO THE PARTNER AND TO THE EXPERIMENTERS' CONFEDERATE

Stimulus person:  Time	Initial Dislike for P				Initial Liking for P			
	Frustrated by the experimenter		Not Frustrated by the experimenter		Frustrated by the experimenter		Not Frustrated by the experimenter	
	P	Confederate	P	Confederate	P	Confederate	P	Confederate
First impression	9.0 <sub>cd</sub>	11.1 <sub>o</sub>	9.5 <sub>d</sub>	11.0 <sub>o</sub>	8.9 <sub>bed</sub>	12.2 <sub>f</sub>	8.4 <sub>bc</sub>	10.9 <sub>o</sub>
End of study	8.4 <sub>bc</sub>	8.2 <sub>b</sub>	6.4 <sub>a</sub>	8.5 <sub>bc</sub>	6.2 <sub>a</sub>	8.7 <sub>bc</sub>	5.8 <sub>a</sub>	6.5 <sub>a</sub>
Change scores	-0.6	-2.9	-3.1	-2.5	-2.7	-3.5	-2.6	-4.4

Note.—The absolute scores obtained at the start and conclusion of the study were subjected to one "repeated measures" analysis of variance. In the above table the higher the mean the more unfriendly the attitude toward the given stimulus person. Cells having the same subscript are not significantly different from each other by Duncan (1955) multiple range test.

by both P and the experimenter still became less unfriendly toward the confederate. Clearly, then, the resentment aroused by the experimenter primarily affected the judgment of the disliked P. It did not interfere with the growth of friendship for the confederate. The previously learned attitude toward the available stimulus person affected the degree to which he was the victim of anger created by someone else. Stimulus objects the subjects had not grown to dislike (i.e., the liked P and the confederate) did not receive anywhere as much generalized aggression.

There was some hostility generalization to the confederate, however. As the table notes, this person was regarded significantly more favorably by those subjects who were not thwarted at all than by the subjects suffering at least one frustration. Nevertheless, these provocations did not retard the development of some friendliness toward the confederate. He may have been associated with the frustrators, but clearly was not one of them.

#### DISCUSSION

The above results, by and large, support the expectations underlying the present experiment. They indicate that the antagonism created by one unpleasant person has a stronger adverse effect on the individual's feelings toward another unpleasant person than on his attitudes toward someone he does not dislike.<sup>4</sup> This enhanced resentment, furthermore, can lead to relatively intense

acts of hostility, as was shown in the preceding experiment in this series (Berkowitz & Holmes, 1960). In that study people given an opportunity to injure a disliked person (by administering electric shocks) in a socially sanctioned manner after having been insulted and criticized by the experimenter, generally performed more of these injurious acts than did the subjects also responding to a disliked person who had not been frustrated by the experimenter or the thwarted subjects given an opportunity to attack someone they presumably liked. Generalizing from these findings to the area of intergroup relations, we can hypothesize that the ethnic groups most likely to become the victim of displaced aggression are those groups the frustrated people had come to regard as being unpleasant for one reason or another (that is, assuming these thwarted individuals interact with these groups or otherwise become aware of them).

These data do not require postulating a stimulus induced generalization of aggressive tendencies from the immediate frustrator to the disliked individual. Many writers (e.g., Freud) have conceived of the aggressive drive as energy continually seeking an outlet. The attacked person supposedly merely provides this opportunity for release. From this point of view, then, we might say the aggressive tendencies aroused, strengthened, or released by the frustrating experimenter were inhibited when the subjects evaluated the neutral confederate or the presumably liked P. The degree of aggression inhibition could have been in direct ratio to the attractiveness of these stimulus objects. Such a formulation, of

<sup>4</sup> It is important to note that these results have been obtained with both male and female college students.



course, is compatible with "balance" (Newcomb, 1959) or "dissonance-avoiding" (Festinger, 1957) propositions. Dissonance (plus aggression-anxiety) would be aroused within a person if he knew he had deliberately injured someone he liked. He would have to suppress any inclinations he might have to hurt the attractive person if he is to avoid this dissonance (and aggression-anxiety).

We agree that this type of phenomenon probably occurs. However, we also believe, perhaps only as an article of faith, that the people the frustrated individual encounters after he is thwarted can serve as stimuli evoking aggressive responses from him. According to our present reasoning the perception of a disliked object is sufficient to elicit such hostility. However, it may be that only people who have inflicted injury on the individual, as was the case in this study, will evoke such generalized aggression, and not every disliked person.

The increased unfriendliness toward the confederate following the frustration of the experimenter can perhaps also be explained as the result of a stimulus generalization process. The provoked subjects could have associated the confederate with the instigator for several reasons. Both were involved in the experiment; both were somewhat unpleasant (although, as noted earlier, the unfavorable attitude toward the confederate might have been held with little conviction); and emotion arousal seems to reduce the use of peripheral cues, resulting in relatively gross discriminations among the available stimulus objects—in essence, flattening and extending the generalization gradient (Easterbrook, 1959). Nevertheless, the generalization of hostility to the confederate in the present study was unexpected and further research is necessary to determine which, if any, of these factors gave rise to the generalization.

The present argument, then, offers a relatively simple solution to the problem of target selection in scapegoating. At least some of the displacement of hostility upon certain minority groups can be accounted for by the thwarted individual's prior dislike for these groups. Feeling this way about them, he presumably associates these people with his most recent frustrators. An industrial worker may

become more aggressive to the Jews in his community after receiving a pay cut because he associates the disliked Jews with the disliked factory owners and managers.

Other variables, however, can also intervene to affect the total strength of the association between ethnic group and immediate frustrator. This linkage may be weakened somewhat by knowledge forcing a discrimination between the particular minority and the frustrating source. It may also be strengthened by additional characteristics shared by the minority and the frustrator. Going back to the illustration of the disgruntled factory worker, suppose he has negative feelings towards both Turks and Jews and encounters a member of each of these groups. His knowledge that Turks generally are not involved in business management may weaken the total associative bond between this ethnic group and the factory owners. Jews, on the other hand, often are businessmen and, more than this, may also be regarded as rich and unscrupulous—just like the owners of the plant. There are several attributes held in common by Jews and the immediate frustrators as far as this individual is concerned, resulting in a heightened association between these people. Conditioned S-R bonds may summate (Hull, 1943, pp. 209–210). Furthermore, as we mentioned earlier, the perceived properties of the disliked minority can lower the individual's internal restraints against aggression. The stereotyped conception of Jews—e.g., that they are grasping and unscrupulous—acts to justify the aggressive inclinations the individual might feel toward this particular group. Believing the stereotype, he need not feel guilty about attacking this group. Knowing the group is in the minority and fairly widely disliked also means it is fairly safe to aggress against Jews. He need not fear retaliatory aggression either from the Jews in the street or from his peers. The consequence of all this is that the Jew is a more probable target for the thwarted worker's hostile tendencies than is the Turk.

#### SUMMARY

Most explanations of social prejudice do not relate the prejudiced individual to the victimized group. They generally are either

"pure drive theories," dealing only with the source of the aggressive tendencies, or "pure stimulus theories," explaining only the characteristics of the attacked minority. The present paper attempts to construct a theoretical bridge between the aggressor and the aggressed-against. The central thesis is that aggression generalizes from the anger instigator to another person in direct proportion to the degree of dislike for this person. Hostile tendencies engendered by frustrations are generalized to those groups whose perceived characteristics result in their being disliked.

In the present study 72 college men were distributed evenly among four experimental conditions created by a  $2 \times 2$  factorial design. Each subject (working in pairs) was first induced to either like or dislike his partner (the P). After this, half of the subjects in each of these conditions were individually frustrated by the experimenter, with the others receiving a more pleasant treatment from him. Then, in the last phase of the study, the two pair members and a neutral peer (actually the experimenter's confederate) were brought together to work on a cooperative task. Each subject's evaluations of his partner (P) and the confederate constituted the dependent variables. The results indicated that the disliked P was the primary victim of the resentment aroused by the frustrating experimenter.

## REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

- ALLPORT, G. W. *The nature of prejudice*. Cambridge, Mass.: Addison-Wesley, 1954.
- ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, 22, 9-39.
- BERKOWITZ, L. Anti-Semitism and the displacement of aggression. *J. abnorm. soc. Psychol.*, 1959, 59, 182-187.
- BERKOWITZ, L., & HOLMES, D. S. The generalization of hostility to disliked objects. *J. Pers.*, 1959, 27, 565-577.
- BERKOWITZ, L., & HOLMES, D. S. A further investigation of hostility generalization to disliked objects. *J. Pers.*, 1960, 28, 427-442.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- DUNCAN, D. Multiple range and multiple *F* tests. *Biometrics*, 1955, 11, 1-45.
- EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, 66, 183-201.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.
- HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
- MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, 43, 155-178.
- NEWCOMB, T. M. Individual systems of orientation. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. III. New York: McGraw-Hill, 1959. Pp. 384-422.
- WHITE, R. K., & LIPPITT, R. *Autocracy and democracy: An experimental inquiry*. New York: Harper, 1960.
- WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychol. Rev.*, 1959, 66, 297-333.
- WILLIAMS, R. M., JR. The reduction of intergroup tensions. *Soc. Sci. Res. Coun. Bull.*, 1947, No. 57.
- ZAWADSKI, B. Limitations of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1948, 43, 127-141.

(Received March 8, 1961)

## CRITIQUE AND NOTES

### BIRTH ORDER AS A SELECTIVE FACTOR AMONG VOLUNTEER SUBJECTS<sup>1</sup>

PAUL C. CAPRA AND JAMES E. DITTES

*Yale University*

Findings (Dittes, 1961; Schachter, 1959) of greater affiliative behavior among first-born persons raise the question of birth order as a selective factor among subjects volunteering for small group experiments. Affiliative and related dependent tendencies may make first borns more vulnerable to the appeal of a recruiter and to the opportunity for participation in small group activity, especially when participation and affiliation appear to be guaranteed by the experimenter. In such contexts, volunteering itself may be understood as an affiliative act.

One hundred Yale freshmen were solicited in their dormitory rooms by a senior student for a small group experiment to be conducted at a later time. The recruiting speech asked freshmen to participate in a small group psychology experiment involving "a group performing a common task cooperatively."

Results appear in Table 1: 36% of first borns and 18% of later borns volunteered for the experiment ( $\chi^2 = 3.8$ ,  $p = .05$ ). Only borns showed an identical rate with other first borns and are therefore included with them in Table 1.

In the hypothetical experiment, had it been conducted, 76% of the subjects would have been first born. This proportion is substantially greater ( $p = .10$ ) than the 61% first borns in the Yale freshman population<sup>2</sup> and close to twice as great as the probable percentage of first borns in the national population, which may be estimated from census data as about 40%.

Review of the birth order information of subjects recruited from freshman dormitories for two other small group experiments (samples of about 100 subjects each) shows first borns to be overrepresented ( $p < .10$ ) among the volunteer

<sup>1</sup> This investigation was supported by Research Grant M-3857 from the National Institute of Mental Health, United States Public Health Service. Part of these results were reported to the American Psychological Association 1961 convention in New York.

<sup>2</sup> The sample of 100 freshmen used in this investigation was representative of the entire freshman class with respect to the proportion of first borns. Of the freshmen found in their rooms, 61% were only- or first born. This is identical with the percentage of only- and first borns found in a 25% sample of the records of the entire Yale freshman class.

TABLE 1  
PROPORTION OF VOLUNTEER SUBJECTS  
BY BIRTH ORDER

Birth order	Volunteered	Declined
First born	22	39
Later born	7	32

subjects, as compared with the total Yale population.

These results specify one sample bias likely in any small group study using volunteers. The common allegation that most current knowledge of small groups represents only the psychology of college freshmen and sophomores could perhaps be more precise by restricting it to the first-born freshmen and sophomores. If our interpretation is correct, this bias would operate similarly in those instances of compulsory participation for course requirements, so long as the student has a choice of participating in one of several experiments. Our interpretation is that the appeal of guaranteed interaction in a small group study is more likely to attract first borns.

The consequences of this bias are most serious when variables are being studied in which first borns are known to differ from later borns. This presumably would include the large class of variables related to affiliation, such as dependence, cohesiveness, attractiveness of the group, and such processes as suggestibility and conformity that are known to be related to the attractiveness of the group. The effects obtained with such variables are likely to be exaggerated among volunteer subjects (meaning a higher proportion of first borns) as compared with results obtained in a random sample of the population. Some .05 levels of significance obtained with these variables among volunteer subjects would not have been reached with a random sample.

#### REFERENCES

- DITTES, J. E. Birth order and vulnerability to differences in acceptance. *Amer. Psychologist*, 1961, 16, 358. (Abstract)  
SCHACHTER, S. *The psychology of affiliation*. Stanford: Stanford Univer. Press, 1959.  
(Early publication received November 2, 1961)



# THE COGNITIVE STRUCTURE OF A SOCIAL STRUCTURE<sup>1</sup>

CLINTON B. DE SOTO AND JOHN J. BOSLEY

Johns Hopkins University

The term *cognitive structure* is doubly distressing to many psychologists. By *cognitive*, it refers to something not very accessible, and by *structure* it teasingly adds that this ghost is nicely organized or articulated. The present article attempts to relieve some of this distress. It reports a novel technique for a relatively comprehensive and bias-free evaluation of the properties of a cognitive structure, together with an application of the technique that appears to vindicate the term.

The technique may be described as the study of the errors subjects make when they learn a mapping of elements into the cognitive structure. For the present study, it was supposed that Johns Hopkins undergraduates have a cognitive structure corresponding to their four classes, freshman, sophomore, junior, and senior, and a sample of undergraduates was required to learn a mapping of people into this structure. Specifically, the subjects were taught freshman, sophomore, junior, and senior as labels for men's names in what could be characterized as a paired-associates verbal learning task with the names as stimuli and the labels as responses.

There are two principal ways of analyzing the subjects' errors in a task like this, and both of them were used here. One is to determine where in the structure the subjects learn fastest, under the supposition that these points are their anchors in the structure; the other is to study in detail the confusions among the various points in the structure, under the supposition that confusions are inversely related to distances between points (Shepard, 1958; Torgerson, 1958).

## METHOD

The experimental method resembled that of traditional paired-associates learning experiments. The subject's task is described well by the instructions given him (the experimenter read the instructions aloud while the subject followed on his own copy):

This is an experiment on learning. Your task will be to learn which of the labels "freshman," "sophomore," "junior," or "senior" is correct for each person in the group whose names are typed on these cards.

<sup>1</sup> This work was supported in part by Grant NSF-G10884 from the National Science Foundation. J. J. Bosley participated as a Public Health Service Research Fellow.

Each card has a man's name on the front, and on the back of the same card is the label which is correct for him. When I give you the cards, they will be arranged so that the names are up. The pace for the experiment will be set by means of the device you see here. This device makes a sound every 3 seconds. Each time it sounds, read aloud the name on the top card, and tell me the one of the labels above that you think correctly goes with that person. Then pick up the card and turn it over, so that you can see whether you were right or wrong, and place the card to one side after you have checked your answer.

There will now be a different card before you on the pile. When the timer sounds again, read aloud the new name and give the label you think is right for it, checking yourself as before and placing the card aside on the new pile. Repeat these steps for each card in the original stack.

After you have gone through all of the cards in this way, I will shuffle them and give them back to you for another run-through. By going through the set time after time, you should eventually learn the proper label for every name in the set.

There were 16 names in the set, 4 with each label. The subjects themselves were also equally divided according to class membership, 7 from each class, totaling 28. Each subject was run individually to a criterion of two successive trials on the entire deck without a mistake. When necessary, a subject was reminded that he had to give an answer for every card.

## RESULTS

The data of interest, the errors made by the subjects en route to learning the correct labels, are summarized in Table 1. This table shows the number of errors averaged over subjects and items as a function of the correct label and the label mistakenly given. It shows, for example, that a name for which the correct label was

TABLE 1  
MEAN NUMBER OF ERRORS

Correct label	Label given by subject				
	Freshman	Sophomore	Junior	Senior	$\Sigma$
Freshman	—	3.48	2.74	1.72	7.94
Sophomore	2.36	—	4.77	3.29	10.42
Junior	2.23	3.54	—	3.29	9.06
Senior	1.78	3.00	4.08	—	8.86
$\Sigma$	6.37	10.02	11.59	8.30	—

freshman was miscalled sophomore a mean of 3.48 times, and was miscalled *something* a mean of 7.94 times.

Inspection of the right-hand column of Table 1 shows that the subjects made the fewest errors on names for which the correct label was freshman, somewhat more when it was senior, slightly more still when it was junior, and the most when it was sophomore. The differences among these means were established as significant at the .025 level by analysis of variance. However, such error rates can be influenced by response sets or guessing tendencies, making it important also to examine the bottom of the table, which shows the frequency with which each label was given incorrectly. It is apparent that freshman was given by mistake very little, senior more often, sophomore still more often, and junior very often. Thus the low frequency with which subjects miscalled freshmen something else is not because of a tendency to guess freshman freely. On the contrary, it is in spite of a tendency to guess freshman infrequently, as demonstrated by the fact that the total for the freshman column is actually less than the total for the freshman row. Similarly, the relatively low frequency with which subjects miscalled seniors something else occurs despite a relatively low frequency of saying senior. The upshot is that the subjects clearly learned the freshmen fastest and the seniors next fastest, with the sophomores and juniors being substantially greater and roughly equal in difficulty.

This conclusion is confirmed by another measure of rate of learning, trials to criterion, which is somewhat less influenced by guessing tendencies. The mean number of trials the subjects required to learn to give the proper label to freshmen (with no subsequent errors) was 13.3; for sophomores, they required 17.5; for juniors, 16.4; and for seniors, 14.6. The differences among these means were established as significant at the .001 level by analysis of variance.

Inspection of the interior cells of Table 1 yields a somewhat different view of the results. The rows of the table show what might be called response generalization gradients. A freshman is miscalled sophomore most, junior less, and senior least. A sophomore is miscalled junior more often than senior. A junior is miscalled sophomore more often than freshman. And a senior is miscalled junior most, sophomore less, and freshman least. This systematic patterning of the errors, evidence in itself of a cognitive structure, encourages an attempt to use the frequencies of errors to determine distances in the structure. There are various ways that this might be done. A very simple

procedure was chosen with the aim of obtaining an ordered metric (Coombs, 1950), or, more precisely, a higher-ordered metric (Siegel, 1956)—a ranking of the distances between the labels, pair by pair. The procedure was to obtain the mean confusions of each pair of labels by adding together the mean errors in both directions, assuming that the fewer the confusions, the greater the distance between the pair of labels. The mean confusions of freshmen and seniors is 3.50, the smallest such figure, indicating that the distance from freshman to senior is the greatest, as it should be. The next largest distance is from freshman to junior (4.97 mean confusions). Third is from freshman to sophomore (5.84 confusions), and a close fourth is from sophomore to senior (6.29 confusions). It might have been expected that the latter two-step distance would be larger than the former single-step distance, but it is not necessary for a consistent higher-ordered metric that it be so. Fifth is from junior to senior (7.37 mean confusions), and sixth and smallest is the distance from sophomore to junior (8.31 mean confusions). This ranking of distances is consistent and indicates that the distance from freshman to sophomore is a very large one, that from sophomore to junior is very small, and that from junior to senior is intermediate, all the labels lying on a single dimension.

One unfortunate characteristic of Table 1 deserves some discussion. The table is not as symmetric as one might like. For instance, freshmen are miscalled sophomore substantially more often than sophomores are miscalled freshman. And these asymmetries are not entirely ascribable to sampling error. According to *t* tests, the difference between opposite entries is significant at the .01 level for two of the six pairs (freshman-sophomore and sophomore-junior). Does this imply that the distance from freshman to sophomore is shorter than the distance from sophomore to freshman? Not necessarily. The response sets or guessing tendencies which were alluded to earlier may play a part. It was suggested that the subjects guessed freshman too little and junior too freely. Various corrections for such response sets are possible. One simple procedure is to add a small constant to the entries in each column so as to bring the column totals into symmetry with the row totals. The necessary constants are .39 for the freshman column (the difference between the total for the freshman column and the total for the freshman row), .10 for the sophomore column, —.63 for the junior column, and .14 for the senior column. It is to be emphasized that the aim of the correction is *not* to bring the column totals into equality—



they should in fact vary if there are generalization gradients—but to make the table more symmetric about the diagonal running from upper left to lower right. The effect of the correction is to reduce the asymmetries markedly, making them all nonsignificant, without changing the ordering of mean confusions arrived at earlier. This outcome tends to sustain the results of the direct unrectified averaging.

The main reason for drawing an equal number of subjects from each class in this experiment was to check on any possible differential "social perspectives" (Hartley & Hartley, 1955; Newcomb, 1950; Sherif & Sherif, 1956) that might accompany class membership—say, a tendency for a subject's own class to be an anchor, or a tendency for the distances to classes neighboring his own to be large relative to the distances between remote classes. No substantial evidence of such differential social perspective emerged. However, one subject provided striking evidence of another kind of personal distortion of the structure. For this subject, freshman to junior seemed to be by far the shortest distance (5.75 mean confusions), with sophomore to junior next shortest (3.50 mean confusions) and freshman to sophomore next (2.62 mean confusions). Senior was very distant from the other labels, having 1.25 mean confusions with freshman, 1.25 with junior, and 1.88 with sophomore. This wildly deviant pattern, at first inexplicable, made more sense after the experimenter asked the subject what class he belonged to (a question routinely asked after the experimental session, as a final check). The subject replied that he was both a freshman and a junior—the university called him a freshman, but his classmates were juniors. He had flunked out two years previously and recently been readmitted. His predicament seems rather faithfully reflected in his confusions during learning. (He was finally assigned to the junior group of subjects on the basis of his statement that his primary allegiance was to the junior class.)

#### DISCUSSION

As a demonstration of a dimension of response generalization in verbal learning, this experiment is probably unmatched in the literature (see Underwood, 1950). The subjects were on the job, so to speak, treating the labels as ordered in a task for which the ordering was irrelevant, at the expense of such a familiar source of response confusion as the homophonic similarity of sophomore and senior. And precisely because the ordering showed up strongly where it was irrelevant, it is seen to be, for the subjects, an essential, unforgettable property of the labels, fundamental in

their cognitive counterpart of the social structure. This conclusion seems in good accord with earlier arguments for the importance of an ordering schema in social cognition (De Soto, 1960, 1961).

*Anchoring.* In like manner, because the task in no way demanded it, the indication that freshman and senior served as anchors attains a special significance. It had seemed fairly likely, in view of what social psychologists have said about people's membership groups serving as anchors, that the subjects would learn most rapidly the names belonging in their own class. Instead, any such tendency disappeared in the race to learn the freshmen and seniors. End-anchoring was the winner of the impartial test.

End-anchoring is a well-known phenomenon in absolute judgments (Volkman, 1951). And, in the sense of better performance at the ends than in the middle of the ordering, end-anchoring is also seen in rote serial learning (Deese, 1958). The present end-effect strongly resembles these earlier end-effects, even to the detail of primacy-finality asymmetry—freshman being learned faster than senior. It would appear that end-anchoring is a universal phenomenon of people's dealings with orderings. Surely one of the first principles of a theory of the cognition of social structures of a theory of the cognition of social structures should be a statement that end-anchoring of social orderings is to be expected.

The concept of anchoring unfortunately lacks refinement—an anchor has been variously defined as a standard or reference point, as a particularly prominent part of a cognitive structure or field, as a part of a cognitive structure around which other parts are built or organized, as a particularly weighty determinant of behavior—but it is probable that these definitions are more complementary than antagonistic.

There are in the literature on social structures various indications of end-anchoring. Sherif, White, and Harvey (1955) found that in newly forming boys' groups the top and bottom statuses were established first. Kahl (1957) reports that people agree best as to the status or prestige of occupations at the extremes of the occupational hierarchy. Homans (1950) discusses a declining New England community in which the social structure was disintegrating and there was no longer much consensus on social standings—except at the top and bottom. The authors have observed that if people are asked what Indian castes they know, they usually mention that the Brahmins are the highest caste and the untouchables the lowest, but are unable to go beyond these first essentials.

These illustrations show that end-anchoring



has the effect of producing consensus among the viewers of a social ordering as to what might be called "reference groups" if that term were not already pre-empted. It is noteworthy that the strength of existing reference group theory (Newcomb, 1950; Sherif & Sherif, 1956) lies in its recognition that reference groups vary, not only from one membership group to another, but also among individuals in a given membership group. The point is made that consensus is lacking where it had been assumed to exist. But this is only one side of the picture and should not direct attention altogether away from the places where consensus *does* exist. It has been argued that the widespread existence of social orderings depends on a consensual expectation of orderings (De Soto, 1960; De Soto & Kuethe, 1959). To that consensual base should be added end-anchoring, which provides reference groups of a kind on which there is consensus, not only among the members of the society but among outsiders too.

End-anchoring seems to occur even when an ordering lacks ends, as is likely to be the case for orderings of people or events in time. For example, people everywhere tend to speak of the founder, the first member, of their family or lineage or clan, although this founder is more likely mythical than not (Fortes, 1953). Talk about evolution usually centers around the first and last representatives of an ordering—the first man vs. modern man (speculation about the "missing link" depends on the assumption of a last ape and first man), the original economic system vs. the ultimate one. And whoever feeds the evident cognitive hunger for end anchors of orderings, supplying the "ultimate" economic system, for example, is capitalizing, wittingly or unwittingly, on a dangerous human weakness.

*Distances.* The interpretation of errors during learning as reflecting distances between the labels seems at first rather divergent from their interpretation as reflecting anchoring, but the two interpretations are not incompatible, and may even have a logical connection. Taylor (in press) concluded that, in visual perception at least, anchors operate to change the psychological distances in their neighborhood. It is possible that freshman to sophomore is the largest distance because freshman is the principal anchor and that junior to senior is the second largest distance because senior is the second anchor.

In any case, the interpretation of the confusions as reflecting distances, although appealing, should be regarded with some caution, if only because of the novelty of the technique. There is very little precedent for scaling psychological distances through confusions during learning

(Shepard, 1958; Torgerson, 1958)—some sort of judgmental task has almost universally been used in such work—and there appears to be no precedent for such an application of verbal learning.

People do sometimes speak of distances in social structures—of big steps and little ones on their social ladders—and this fact lends a face validity to a distance interpretation of confusions in the present data. And to the authors at least, the distances reported for the undergraduate hierarchy have substantial face validity. Unfortunately, casual interviews of a few subjects failed to provide clear independent confirmation of the experimentally determined distances. But then this failure may demonstrate only the superiority of the confusions technique, which eliminates any problems of the subjects' interpretations of questions about distances by not asking them any such questions. Indeed, the case of the freshman-junior subject suggests that the technique has possibilities as a rather sensitive and unusually neutral and subtle diagnostic instrument. But a final evaluation of the distances obtained in this study must await the results of future applications of the technique.

#### SUMMARY

A paired-associates learning experiment was performed in which men's names were the stimuli and the labels freshman, sophomore, junior, and senior were the responses.

The subjects learned most rapidly to apply the labels freshman and senior correctly, a result that was interpreted as end-anchoring. The errors the subjects made during learning were patterned in such a way as to demonstrate generalization gradients for each label. On the supposition that the frequency of confusions of labels was inversely related to the psychological distance between them, it was determined that psychologically the labels fell on a single dimension, and that the distance from freshman to sophomore was largest, the distance from sophomore to junior, shortest, and the distance from junior to senior of intermediate size.

Some discussion was given of the broader social significance of the results, especially of the end-anchoring.

#### REFERENCES

- COOMBS, C. H. Psychological scaling without a unit of measurement. *Psychol. Rev.*, 1950, 57, 145-158.
- DEESE, J. *The psychology of learning*. (2nd ed.) New York: McGraw-Hill, 1958.
- DE SOTO, C. B. Learning a social structure. *J. abnorm. soc. Psychol.*, 1960, 60, 417-421.
- DE SOTO, C. B. The predilection for single orderings. *J. abnorm. soc. Psychol.*, 1961, 62, 16-23.

- DE SOTO, C. B., & KUETHE, J. L. Subjective probabilities of interpersonal relationships. *J. abnorm. soc. Psychol.*, 1959, 59, 290-294.
- FORTES, M. The structure of unilineal descent groups. *Amer. Anthropologist*, 1953, 55, 17-41.
- HARTLEY, E. L., & HARTLEY, R. E. *Fundamentals of social psychology*. New York: Knopf, 1955.
- HOMANS, G. C. *The human group*. New York: Harcourt, Brace, 1950.
- KAHL, J. A. *The American class structure*. New York: Rinehart, 1957.
- NEWCOMB, T. M. *Social psychology*. New York: Dryden, 1950.
- SHEPARD, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, 55, 509-523.
- SHERIF, M., & SHERIF, C. W. *An outline of social psychology*. (Rev. ed.) New York: Harper, 1956.
- SHERIF, M., WHITE, B. J., & HARVEY, O. J. Status in experimentally produced groups. *Amer. J. Sociol.*, 1955, 60, 370-379.
- SIEGEL, S. A method for obtaining an ordered metric scale. *Psychometrika*, 1956, 21, 207-216.
- TAYLOR, M. M. The effect of anchoring and distance perception on the reproduction of forms. *Percept. mot. Skills*, in press.
- TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.
- UNDERWOOD, B. J., & HUGHES, R. H. Gradients of generalized verbal responses. *Amer. J. Psychol.*, 1950, 43, 422-430.
- VOLKMANN, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951. Pp. 273-294.

(Received February 14, 1961)

*Journal of Abnormal and Social Psychology*  
1962, Vol. 64, No. 4, 307-312

## EFFECTS OF MUSIC UPON GSR OF DEPRESSIVES AND SCHIZOPHRENICS<sup>1</sup>

EDWARD W. WEIDENFELLER AND GEORGE H. ZIMNY

Marquette University

The purpose of the present study is to investigate the effects of two pieces of music lying at separate points along a single exciting-calming dimension upon GSR (electrodermal response) of depressive and schizophrenic patients.

The effects of music upon GSR in normals have been studied by Dreher (1947), Henkin (1957), and Traxel and Wrede (1959). More recently, in an as yet unpublished study by the present authors, the effects of exciting, neutral, and calming music upon GSR and heart rate in normals were investigated. Many factors such as pieces of music, experimenter, and methods of stimulation present in that study are the same as those used in the study reported here. It was found in the previous study that the exciting music produced a decrease, the calming music an increase, and the neutral music no change in electrical resistance.

There does not appear to be any study on the specific question of the effects of music on GSR of psychotics. However, GSR has been used as a

response measure with psychotics, and the influence of music upon some responses of psychotics has been studied. With reference to GSR, Paintal (1951) found differences between normals and psychotics under threat of electrical shock but no difference in the presence of shock. Herr and Kobler (1953) found differences in GSR between neurotics and normals, and Enste and Meyer (1953) found GSR differences between various types of psychosis and between psychotics and normals. With reference to music, Gilman and Paperte (1949) investigated its effects upon behavioral changes in psychotic and nonpsychotic groups. They found that exciting music had no differential effect upon the two groups whereas the calming music had a greater effect in the expected direction on the psychotic group than on the normal group. Simon, Holzberg, Alessi, and Garrity (1951) played eight piano pieces to normals, schizophrenics, manics, and psychotic depressives and asked them to indicate whether each piece was happy, sad, or neither and whether they liked or disliked each piece. They found no significant differences between the normals and the psychotics in their identification of the mood of the music. Skelly and Haslerud (1952) played a number of short pieces of music to 39 female apathetic schizophrenic patients. The presenta-

<sup>1</sup> This study was carried out with the cooperation of the administration and staff of the Milwaukee County Hospital for Mental Diseases. The authors wish particularly to thank Chris Buscaglia, Medical Director, John Liccione, Chief Psychologist, and Leo Muskatevc, Director of Musical Therapy.



tion of the music to the patients began with pieces rated as depressing, continued with a gradual shift to pieces rated as exciting, and concluded with a shift back to pieces rated as depressing. The patients were given such a series of pieces lasting 20-30 minutes daily for over a month. It was found that in comparison to their base activity level the patient showed a statistically significant increase in activity level during the playing of the exciting music. No carry-over effect of the exciting music on activity level after a 6-hour period was found, thus indicating that the stimulating effects of the music were short-term.

The present study is composed of two separate but related experiments. Experiment I employs depressive patients and Experiment II employs schizophrenic patients. The hypothesis being tested in each experiment is that the music produces differential effects upon GSR and, more specifically, that the piece of music designated as exciting produces a decrease in electrical resistance of the skin and the piece of music designated as calming produces an increase in electrical resistance of the skin. A change in resistance is considered as a departure from the level of resistance present at the start of the music.

The assumption is made here that GSR is a manifestation of emotional response. A decrease in electrical resistance is thus interpreted as being due to an increase in emotional excitation while an increase in electrical skin resistance is interpreted as being due to a decrease in emotional excitation. The present study using GSR is thus conceived of as an investigation of the effect of music upon the emotional level of psychotics.

## METHOD

### Subjects

In Experiment I, subjects were 14 female and 4 male patients who were diagnosed as having one of the following types of depressed illness: involutional, neurotic, manic-depressive depressed, or reactive. The mean age was 52 years. The 18 subjects were randomly selected from the total available hospital population of 45 such patients. Experimental data on 8 additional subjects were discarded, since within one week after being run they were judged by two clinical psychologists on the basis of individual interviews as no longer clinically depressed.

In Experiment II, subjects were 13 female and 5 male patients diagnosed as possessing one of the major types of schizophrenic illness. Patients who presented a history of any different diagnosis were excluded. The mean age was 48 years. The 18 subjects were randomly selected from the total available hospital population of 101 older patients who had been admitted in the previous 3 years.

### Apparatus and Material

Selected 6-minute portions of two musical pieces, Dvořák's "Final Movement" of the *New World Symphony* and Bach's "Air for the G String" (ML5115 and ML5065, Columbia Record Company) were judged by the authors to be exciting and calming, respectively. As an empirical check on the authors' judgment, 59 lower division undergraduates were asked to rate each (edited) piece on a five-point exciting-calming scale and on a five-point familiarity scale. Comparisons of the ratings by means of Wilcoxon's matched-pairs signed-ranks test (Siegel, 1956) showed a significant difference ( $p < .001$ ) between the two pieces in the expected direction on the exciting-calming scale. In addition, the "Air for the G String" was rated as the more familiar piece ( $p < .001$ ). It does not seem likely, however, that the difference in familiarity found for the college students would also be found for the psychotics used in this study.

A tape of the selected portion of each piece was played on a Wollensak tape recorder over a high fidelity loudspeaker. GSR was measured using a psychogalvanometer (Stoelting, Model No. 24207). The measures consisted of reaction units read from a dial located on the front of the galvanometer. In the present experiment, a change of 12.5 reaction units from the zero dial reading represented a change in resistance of 1000 ohms. This galvanometer does not permit the determination of a base level. Before each of the two pieces of music was played to each subject, the galvanometer was adjusted to match the (unknown) level of resistance of the subject, resulting in a dial reading of zero.

### Procedure

In both experiments, each subject was run individually and was exposed to both musical pieces. Two sequences of pieces were used, namely, AB and BA, with each subject assigned randomly to a sequence. The subject was seated in a chair facing a wall and surrounded on the other three sides by hospital bed screens. The experimenter and all recording and playing apparatus except the loudspeaker were located outside the screened area (about 7' x 5' in size).

The sequence of events for each subject was as follows: instructions to the subject and connection of the galvanometer to the subject (3 minutes); pre-music measurement (1 minute); measurement with music playing (6 minutes); postmusic measurement (1 minute); rest period outside the screened area (4 minutes). The entire sequence was then repeated with the second piece of music beginning with the 3-minute instruction and connection period. The subjects were instructed to sit quietly and merely listen to the music. The number of reaction units shown on the galvanometer were recorded every 15 seconds, yielding a total of 24 readings per subject for each piece.



## RESULTS

## Experiment 1

The results for the exciting and calming music played to the depressive patients, expressed in mean GSR reaction units, are summarized in Table 1. The mean values shown for each successive minute of time were derived by averaging the four readings obtained during any given minute for a given subject and then computing a grand mean for all 18 subjects.

The main hypothesis of the experiment is that the exciting music produces a decrease and the calming music produces an increase in the electrical resistance of the skin. An increase or decrease is a departure from the level of resistance present at the start of the music, a level designated by zero reaction units. As shown in Table 1, the means for the exciting music are all positive (indicating a decrease in resistance) and those for the calming music are all negative (indicating an increase in resistance). The significance of the difference of the mean number of GSR reaction units from a mean of zero reaction units was tested using the  $t$  test for each period of elapsed time for each piece of music. As shown in Table 1, all the  $t$  values are significant ( $p < .005$  or  $p < .0005$ ) except the one for the first minute of calming music. The hypothesis is thus confirmed when the exciting music is played for 1-6 minutes and is confirmed when the calming music is played for 2-6 minutes.

Examination of the mean values at each time period indicates that after 1 minute the exciting music produced a decrease in electrical resist-

ance. After 2 minutes, an even greater decrease occurred, but from then on the decrease in resistance began to approach a plateau. How long this plateau lasts and in what direction subsequent responses tend are questions needing investigation. For the calming music, an increase in electrical resistance occurred after 2 minutes. The level of resistance remained substantially the same for the remaining time periods. The magnitude of change in resistance, disregarding whether the change was in the direction of an increase or decrease from zero, was much greater in response to the exciting than to the calming music. Tests of the significance of the differences between the two musical pieces in the magnitude of change from zero, disregarding direction of change, yielded significant  $t$  values (minimum  $p < .02$ ) at 2 through 6 minutes. Thus the response of the depressives to each piece of music differed in direction (decrease vs. increase), latency (1 vs. 2 minutes) and magnitude (large vs. small).

A comparison between the results obtained for each piece of music shows that for the one-minute period the  $t$  value for the difference in variances (Walker & Lev, 1953) is not significant while the  $t$  value for the difference in means is significant ( $p < .005$ , one-tailed test). For each of the remaining five time periods, the  $t$  values for differences in variances and in means are significant ( $p < .0005$ ). These analyses clearly demonstrate that the playing of the exciting and calming music produced a difference in the level of electrical resistance of the skin (relative to zero) and in

TABLE 1  
MEANS, STANDARD DEVIATIONS, AND  $t$  VALUES\* OF CHANGES IN GSR REACTION UNITS FOR  
EIGHTEEN DEPRESSIVES

Music	Measure	Elapsed time in minutes					
		1	2	3	4	5	6
Exciting	$M$	4.81	11.19	13.33	13.94	15.94	16.43
	$SD$	5.32	10.63	12.65	13.04	12.88	16.48
	$t_{\bar{x}-0}$	3.73**	4.34***	4.34***	4.41***	5.09***	4.11***
Calming	$M$	-.23	-3.94	-3.85	-3.43	-4.47	-4.21
	$SD$	4.63	2.96	3.14	3.59	3.18	4.31
	$t_{\bar{x}-0}$	.29	5.47***	5.06***	3.94**	5.80***	4.01***

\* Mean changes tested against a hypothesized population mean of zero ( $H:\mu = 0$ ). All tests are one-tailed.  
 \*\*  $p < .005$ .  
 \*\*\*  $p < .0005$ .

TABLE 2

MEANS, STANDARD DEVIATIONS, AND *t* VALUES<sup>a</sup> OF CHANGES IN GSR REACTION UNITS FOR EIGHTEEN SCHIZOPHRENICS

Music	Measure	Elapsed time in minutes					
		1	2	3	4	5	6
Exciting	<i>M</i>	4.93	8.88	12.22	14.01	13.47	16.42
	<i>SD</i>	5.95	7.33	9.96	11.92	12.29	10.54
	$\frac{t}{N-0}$	3.42**	4.99***	5.05***	4.85***	4.52***	6.41***
Calming	<i>M</i>	-.15	-2.33	-3.04	-2.99	-3.97	-4.10
	<i>SD</i>	3.26	2.99	4.31	4.62	4.28	5.10
	$\frac{t}{N-0}$	.19	3.19**	2.89*	2.67*	3.82**	3.31**

<sup>a</sup> Mean changes tested against a hypothesized population mean of zero ( $H: \mu = 0$ ). All tests are one-tailed.\*  $p < .01$ .\*\*  $p < .005$ .\*\*\*  $p < .0005$ .

the consistency of the level of resistance. The response of the depressives to the exciting music was much less consistent than was their response to the calming music.

### Experiment II

The results for the exciting and calming music played to the schizophrenic patients are summarized in Table 2. The mean value for each time period in Table 2 was computed in the same manner as in Experiment I. The main hypothesis of the experiment, namely, that the exciting music produces a decrease and the calming music produces an increase in the electrical resistance of the skin, was tested by determining the significance of the difference of the mean number of GSR reaction units from a mean of zero reaction units. As shown in Table 2, all the *t* values are significant (minimum  $p < .01$ ) except, again, for the first minute of the calming music. The hypothesis is thus confirmed for the exciting music and, with one exception, for the calming music.

As shown in Table 2, a significant decrease in electrical resistance occurred in response to the exciting music after 1 minute, and this decrease in resistance tended to become considerably greater over time. For the calming music, a significant increase in resistance took place after 2 minutes, but the level of resistance tended to remain the same over time. The magnitude of change in resistance from zero, disregarding the direction of change, was significantly greater (using the *t* test,  $p < .01$  at 2 through 6 minutes) for the exciting than for the calming mu-

sic. Thus, as was found for the depressives, the response of the schizophrenics to the two pieces of music differed in direction, latency, and magnitude.

A comparison between the results for the two pieces of music shows that all the *t* values for the differences in variances are significant (minimum  $p < .01$ ) as are all of the *t* values for the differences in means ( $p < .0005$ , one-tailed test). For all six periods of time, then, the playing of the exciting and calming music produced a difference in the level of electrical resistance of the skin (relative to zero) and in the consistency of the level of resistance. As was found for the depressives, the response of the schizophrenics to the exciting music was considerably less consistent than their response to the calming music.

### DISCUSSION

The major finding of the two experiments is that the pieces of music judged by college students to be exciting and calming were capable of producing differential changes in GSR of psychotics. These same pieces of calming and exciting music have also been found by the authors to produce similar changes in GSR of normals. It seems, then, that the effects of these musical stimuli are quite general.

A more detailed comparison between the psychotics used in this study and the normals (college students) used in the previous study indicates that the exciting music produced a greater decrease in the electrical resistance of the normals than of the psychotics while variability was

similar for both groups. This comparatively muted response of the psychotics to the exciting music might be explained by their reduced emotional contact with reality, particularly reality capable of arousing them. The response of the normals and the psychotics to the calming music was similar with respect to the magnitude of increase in electrical resistance, but the normals tended to be somewhat more variable in their response.

In the present study, the most pronounced difference between the results for the depressives and the schizophrenics lies in the greater variability of the response of the depressives to the exciting music. The conception of the GSR as a manifestation of an emotional response of the subjects to the music provides the basis for a possible interpretation of this difference in variability. It is assumed here that the depressive is characterized by a reduction of response to those emotional stimuli that are not related to the affective or mood component of his depression. Normal efforts to cheer up the depressive, for example, are generally unsuccessful. It is also assumed that the exciting music is not related, on the whole, to the affective component of the depression of the patients used as subjects. The music, however, is capable of arousing a response, as has been shown in this and the previous study by the authors. Observations of the depressives during the playing of the exciting music gave the impression that some were trying to resist being influenced by the stimulus. The implication of these observations is that the depressives were susceptible to influence, but did not want to be influenced, and therefore made an effort to resist being influenced. The greater variability of response by the depressives to the exciting music might well be due to a variation in success with which each patient was able to resist being emotionally aroused by the exciting music. The schizophrenics, on the other hand, were shown to be more consistent in their response to the exciting music.

The finding in the present study that the calming and the exciting music produced predicted changes in GSR serves to substantiate the findings for exciting and calming music obtained by other investigators (Gilman & Paperte, 1949; Skelly & Haslerud, 1952) using other response measures. These few studies do lend some limited support to the validity of the use of music with psychotics. The present study indicates the feasibility of at least temporarily modifying the general emotional level of depressive and schizophrenic patients.

## SUMMARY

Two experiments, one using depressive and the other using schizophrenic patients, were conducted to test the hypothesis that calming music produces an increase and exciting music a decrease in electrical resistance of the skin (GSR). In both experiments, a musical piece judged by college students to be exciting and another piece judged to be calming were played for 6 minutes in counterbalanced order to 18 randomly selected depressives and to 18 randomly selected schizophrenic patients. Measures of GSR were obtained for each one of the 6 minutes during which the music was played. The hypothesis was confirmed in each experiment.

It was found for both the depressives and the schizophrenics that the decrease in electrical resistance due to the exciting music was of greater magnitude and shorter latency than the increase in resistance due to the calming music. Comparison of the results for the two pieces of music within each experiment demonstrated a difference in the level of electrical resistance due to the music and in the consistency of the level of resistance. The response to the exciting music was less consistent than the response to the calming music. The changes in electrical resistance are interpreted as due to emotional effects produced by the music. The possibility is thus presented that music can be used to modify temporarily the general emotional level of depressive and schizophrenic patients.

## REFERENCES

- DREHER, R. E. The relationship between verbal reports and the galvanic skin response. Unpublished doctoral dissertation, Indiana University, 1947.
- ENSTE, W., & MEYER, W. Ergebnisse Elektrodermatographischer Untersuchungen an Schizophrenen. [Results of electrodermatographic investigations on schizophrenia.] *Nervenarzt*, 1953, 24, 300-302.
- GILMAN, L., & PAPERTE, F. Music as a psychotherapeutic agent. *J. clin. Psychopathol.*, 1949, 10, 286-303.
- HENKIN, R. E. The prediction of behavior response patterns to music. *J. Psychol.*, 1957, 44, 111-127.
- HERR, V., & KOBLER, F. J. A psychogalvanometric test for neuroticism. *J. abnorm. soc. Psychol.*, 1953, 48, 410-416.
- PAINTAL, A. S. A comparison of the galvanic skin responses of normals and psychotics. *J. exp. Psychol.*, 1951, 41, 425-428.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- SIMON, B., HOLZBERG, J. D., ALESSI, S. L., & GARRY, D. A. The recognition and acceptance of mood in music by psychotic patients. *J. nerv. ment. Dis.*, 1951, 114, 66-78.



- SKELLY, C. G., & HASLERUD, G. M. Music and the general activity of apathetic schizophrenics. *J. abnorm. soc. Psychol.*, 1952, 47, 188-192.
- TRAXEL, W., & WREDE, G. Hautwiderstandsänderungen bei Musikdarbietung. [Changes in psychogalvanic skin response as affected by musical se-

lections.] *Z. exp. angew. Psychol.*, 1959, 6, 293-309.

WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received February 27, 1961)

*Journal of Abnormal and Social Psychology*  
1962, Vol. 64, No. 4, 312-316

## THE EFFECT OF NEGATIVE VERBAL CUES UPON VERBAL BEHAVIOR<sup>1</sup>

JACK SANDLER<sup>2</sup>

*Florida State University*

Within the last 5 years, a number of investigators have successfully expanded the operant conditioning paradigm to human activity, especially with regard to verbal behavior. Since the initial observation by Greenspoon (1955) that verbal behavior can be manipulated by operant techniques, a host of subsequent studies have provided further evidence of the degree to which this phenomenon is subject to the same variables underlying the bar pressing response in rats (Krasner, 1958).

The majority of these studies, however, has been restricted to the use of positive secondary reinforcement as the independent variable. The cues have ranged anywhere from "Mmhm" and "Good" to head nodding, and "paying attention" (Adams & Hoffman, 1960). Although there is good reason to believe that negative reinforcement also has relevancy in this area, this problem had not been systematically attacked. Salzinger (1959) indicated that here, in particular, experimental knowledge is lacking. The value of conducting such an investigation would provide further knowledge regarding the degree to which operant principles could be generalized to human behavior, as well as indicating possible practical implications.

For this purpose, two particular questions were formulated: Does a negative verbal cue have an

effect upon the amount of verbalizations in an operant situation? Do different schedules of the negative cue have different effects?

### METHOD

#### *Subjects*

Sixty male patients selected from a Veterans Administration neuropsychiatric hospital were used in this study. They were all residents of the acute treatment wards and ranged from 28 to 42 years of age. The actual selection employed involved a perusal of the records of all patients on each ward, eliminating those with below average IQ. The remaining men were then individually requested to participate in a research project. The only requirements regarding their selection were average intelligence (IQ of 90) as determined by the Shipley (1940) Institute of Living Scale and cooperativeness. The latter was operationally defined by the procedure as outlined below.

#### *Procedure*

Each subject was conducted individually into the testing room and requested to take a seat facing the desk. The experimenter took the seat behind him thus placing himself outside the visual range of the subject. A recorder and microphone in full view of the subject were then put into operation.

The response class employed in this study was verbalizations in general. In order to provide the means for the manipulation of such an operant response without the subject's awareness of the nature of the problem (Adams, 1957), the experiment was disguised as an attempt to investigate the validity of the TAT (Murray, 1943) and the Symonds' (1949) Picture Story Test. To provide for this the following instructions were then given:

I am interested in finding out something about a test. These cards consist of different pictures. It's supposed to be very easy for anyone to interpret the important point in every picture. I would like to know just how easy this really is. I'll show

<sup>1</sup> This paper is based on a dissertation submitted to the Graduate School of Florida State University in partial fulfillment of the requirements for the PhD degree. The writer wishes to express his sincere appreciation to Joel Greenspoon, dissertation chairman, and to the staff of the Veterans Administration Hospital, Gulfport, Mississippi, without whose generous cooperation this study could not have been undertaken.

<sup>2</sup> Now at the Veterans Administration Hospital, Coral Gables, Florida.

TABLE 1

MEANS, *SD*s, AND *F* RATIOS OF NUMBER OF RESPONSES  
PRODUCED PER THIRTY-SECOND INTERVAL FOR  
ALL THREE PHASES

Group	Phase I		Phase II		Phase III	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	41.87	12.03	31.63	11.51	39.69	12.84
2	37.49	12.77	32.14	12.73	35.82	14.83
3	40.27	6.85	24.19	12.37	29.88	10.76
4	34.86	8.38	35.30	8.96	36.24	10.67
<i>F</i> ratio	1.28		16.78*		1.51	

\* Significant at or beyond .01 level.

number of words produced per 30-second period.

Table 1 presents the mean response rate for each group throughout the three phases. The experimental design provided for an evaluation of the data along two lines. Between-group performances for each phase constituted one analysis; within-group performances from phase to phase constituted another analysis.

#### *Between-Group Comparisons*

Despite the use of a random procedure, Table 1 reveals considerable differences in initial mean response rate from a low of about 35 words per 30-second period for Group 4 to a high of about 42 words per 30-second period for Group 1. However, the results of an analysis of variance indicated in Table 1 reveal that these differences were not significant. On the other hand, the analysis of variance *F* ratio for the training phase data was significant beyond the .01 level. Furthermore, the relative positions of the four groups had changed as compared to their original levels. In Phase II, the control group revealed the highest level of verbalizations, with the variable-interval group responding at a mean rate far below that of the other three groups.

During Phase III, the differences in mean response rate diminished and this was reflected in the nonsignificant *F* ratio revealed in Table 1. The three experimental group rates drew close together and shifted upward once again. As in Phase I, prior to the introduction of the negative cue, subjects in Group 1 revealed the highest mean response rate. Subjects in Group 3, however, continued to produce words at a lower rate than the other three groups.

While there was evidence to assume that the differences revealed during the training phase could be attributed to the various schedules of the negative cue, several possibly confounding variables had to be isolated. One of these, initial

these pictures to you one at a time. In each case, what I want you to do is tell me what you think is happening. I am especially interested in what you think the people are feeling, what events led up to the present situation, how things will turn out, and so forth. Try to use your imagination as best as you can. In order to let you know how you're doing, I'll tell you when you're off the track. In other words, I won't say anything unless you start missing the boat; but regardless of whether or not I say anything, keep talking until you're satisfied that you can think of nothing more to say about a picture. Then, lay the card down on the table and I'll hand you the next one.

The combined cards of the two tests were randomized and were presented to each subject individually. Recordings were kept of the time and number of cues administered. The entire experimental procedure was divided into three continuous phases. Phase I lasted 4 minutes and constituted an attempt to determine operant rate of verbal production. Since to determine conditioning requires the presence of an existing response, those subjects whose initial latency was over 20 seconds, or whose initial verbalizations were less than 10 seconds, were discarded as "uncooperative." On the basis of this criterion, 11 subjects were discarded.

Phase II constituted the Training Phase. The negative cue employed in this study was "Unh unh." Although arbitrarily selected there was strong a priori reason to assume its influence as a negative generalized reinforcer as suggested by Skinner (1953). Greenspoon (1955) demonstrated that it was effective for at least one response class. This cue was administered by means of differential interval schedules. Subjects in Group 1 received the cue every 20 seconds, (FI: 20 seconds), in Group 2 every 40 seconds (FI: 40 seconds), and in Group 3 in a variable interval order. Here the cue occurred anywhere from 5 to 35 seconds around a mean of 20 seconds. Thus the schedules in Groups 1 and 2 represent periodic reinforcement, and in Group 3 aperiodic reinforcement. Subjects in Group 4 (the control group) received no cues during this time but were permitted to verbalize for the same length of time as subjects in Group 1. Phase II lasted until 20 such cues had been administered for all groups.

The last part of the experiment (Phase III) involved the withdrawal of the cue and provided an opportunity for all subjects to verbalize for an additional 4 minutes.

At the conclusion of the procedure, each subject was asked if he thought the experimenter had in any way made him modify his behavior and if he could think of any way that his verbal rate had been changed. The design was thus constructed to enable each subject to act as his own control as well as to provide the means for a comparison between three experimental groups and a control group.

#### RESULTS

In order to analyze the data systematically, the response measure was organized in terms of

verbal facility, was adjusted for by means of an analysis of covariance (Edwards, 1956). Table 2 presents the results of this analysis. The obtained *F* value was significant beyond the .01 level indicating that the training phase differences were not attributable to initial differences in verbal facility.

An additional possibly confounding variable was the variation in number of 30-second periods experienced by the four groups. The covariance *F* ratio (Edwards, 1956, p. 349) for these data was significant beyond the .01 level indicating that the Phase II differences were not attributable to differences in length of time in the training phase.

The Phase III data provided some interesting information despite the lack of significance between the group differences. Using operant level performance (Phase I) and training level performance (Phase II) as predictor variables (Edwards, 1956) the Phase III data were subjected to analyses of covariance. The former *F* ratio was not significant revealing little response change in relative positions between the four groups. The latter analysis, however, was significant far beyond the .01 level indicating that the groups made a considerable change in their relative positions between Phases II and III.

*Within-Group Comparisons*

Inspection of Table 1 reveals that the control group maintained a relatively stable mean rate of response, fluctuating less than 2 words per 30-second period from phase to phase. The experimental groups, on the other hand, demonstrated extensive changes in performance from phase to phase. Considering the differences between Phase

I and Phase II, Group 1 response rate dropped about 10 words per 30-second period, Group 2 level decreased by about 5 words per 30-second period, while subjects in Group 3 talked an average of about 16 words less per 30-second period.

With the cue withdrawn (Phase III) all of the experimental groups increased their verbal production rate when compared with Phase II performances. Actually, Groups 1 and 2 almost returned to their operant level strength. Subjects in Group 3, however, were still considerably below their initial response level, although also reflecting an increase over Phase II production level.

None of the subjects was capable of verbalizing any relationship between the experimenter's cues and his own performance, nor could any express any recognized change in his rate of verbal production.

DISCUSSION

In terms of answering the two questions which generated the study, there is strong evidence to indicate that the cue, "Unh unh," acted as a negative reinforcer (Skinner, 1953) by increasing response probability when removed from a conditioning situation. In addition, it also appears likely that the different schedules of the negative cue had different effects.

With respect to the first finding, it is easily seen that the present experiment confirms much of the data derived from operant studies on sub-human organisms. It is generally recognized, for example, that presenting an aversive stimulus to an operant response will have two effects (Estes, 1944; Jenkins & Stanley, 1950; Skinner, 1953): a decrease in response probability when adminis-

TABLE 2  
ANALYSES OF COVARIANCE OF PHASES II AND III

Adjusted variable	Source	SS	df	F
Phase II (Adjusted for initial differences)	Between-groups	2,741,008.38	3	36.61*
	Within-groups	1,372,561.24	55	
	Total	4,113,569.62	58	
Phase II (Adjusted for time periods)	Between-groups	1,551,829.47	3	33.05*
	Within-groups	671,686.64	41	
	Total	2,123,516.11	44	
Phase III (Adjusted for initial differences)	Between-groups	63,368.63	3	4.18
	Within-groups	286,511.45	55	
	Total	351,880.08	58	
Phase III (Adjusted for training differences)	Between-groups	335,818.29	3	115.83*
	Within-groups	53,151.43	55	
	Total	388,969.72	58	

\* Significant at .01 level.



tered, or suppression of the response; an increase in response probability when withdrawn. This is clearly demonstrated in the present study where response rates for the experimental groups diminished during training, but increased when the cue was withdrawn during the final phase, and, in the cases of Groups 1 and 2, almost returned to operant levels.

The most extensive investigation of this phenomenon has been reported by Estes (1944). Estes' conclusions were derived from a series of studies evaluating the effect of electric shock upon the bar pressing response. Both his and the current study, however, seem to fall into the type of punishment situation which involves the presentation of a negative reinforcement. In agreement with Skinner (1953), and as in the present situation, he found that response strength increased when the cue was withdrawn.

While these observations have been confirmed numerous times, the theoretical implications of this phenomenon have been largely neglected. Recently, however, Dinsmoor (1954) formulated a general framework within which free response aversive conditioning could be explained by means of an "avoidance hypothesis." Essentially, this principle implies that the suppressive action of an aversive stimulus is due to the conditioning of avoidance reactions which conflict with the original behavior being modified. Such an explanation might fit the findings of the present study. In Phase II, for example, it is possible that silence, in conflicting with the chain of reactions leading to the response associated with the administration of the negative cue, changed the situation from an aversive one to a non-aversive one. Inasmuch as the experimental arrangement was probably only mildly threatening, the response was not completely suppressed in favor of total silence. Consequently, subjects merely talked at a slower rate and when the cue was completely withdrawn in Phase III, verbal rates began to accelerate, depending upon the schedule to which the subject had been exposed. One can only speculate, however, regarding the outcome of a similar experiment wherein a primary negative reinforcer (such as shock) would be employed as the negative cue. This would provide an additional test of the "avoidance hypothesis."

With respect to the question of schedule influence, here again, a considerable mass of data has been accumulated, primarily on subhuman organisms. In their review of the literature, Jenkins and Stanley (1950) indicate that a continuous schedule is almost always more effective in training than is a partial schedule, but partialing

usually results in greater durability of effect. These principles, however, are adduced largely from studies employing positive reinforcement. Estes' (1944) data suggest that these principles are also true when the independent variable is aversive in nature.

The present study offers no opportunity to evaluate the difference between continuous and partial schedules. However, it is reasonable to assume that cue administration in Group 2 (every other 20 seconds) represents greater partialing than in Groups 1 and 3. Since the average drop in response rate from Phase I to Phase II for Group 1 was twice that of Group 2, some confirmation of Estes' (1944) findings is revealed. This is further substantiated by comparing the variable-interval group performance with that of Group 2. Clearly, verbal rate decreased most markedly under this condition. Thus these data support the observation that the shorter the interval between cue administration, regardless of whether or not the cue is aversive in nature, the greater the effects in terms of training.

With respect to durability of effect, it is obvious that the variable interval schedule was again the most effective. This appears to be due largely to the fact that response rate reduction for this group during Phase II was so great, these subjects simply did not have enough time during the final stage of the experimental stage to return to their initial level of response strength. Thus the durability of the effect apparently relates directly to the differences in Phase II.

There seems to be little question that the reinforcement "Unh unh," presented on a variable interval basis, represented a more effective aversive stimulus than the same cue administered periodically (fixed interval). It is difficult to determine why this particular arrangement should have had such a profound effect. There is little previous research with which to relate these findings, particularly in the field of punishment of a "free responding" operant. Azrin (1956) reports a study which investigated the effects of fixed and variable interval as well as immediate and nonimmediate punishment on the pigeon pecking response. Although Azrin's focus was primarily in terms of evaluating the effects of shock associated with response and shock not associated with response, his findings have some bearing on the present study. Regarding the two conditions which have the most relevance for this investigation (fixed interval shock without response-shock correlation and variable interval shock without response-shock correlation) under the first condition, Azrin found a negatively accelerated pattern of response prior to the delivery of

each shock, followed by recovery. In the second procedure, response rate was more erratic with, however, no consistent deviation from a uniform rate of responding as in the first case. Furthermore, a comparison of the cumulative curves as well as the statistical data which he presents reveals a lower overall rate of response for the aperiodic conditions than for the periodic conditions.

These findings, then, are quite similar to those of the present study which revealed the greater effectiveness of the aperiodic schedule and reflects the possibility that the relative influence of schedules under aversive stimuli seems to be opposite to the effects of positive reinforcement presented under the same conditions (Ferster & Skinner, 1957). That is, the effects of periodic and aperiodic punishment, whether primary or secondary, are analogous to the effects of periodic and aperiodic reward. The difference seems to be one of direction.

With regard to further research in this area, this study has demonstrated that certain operant principles can be generalized to verbal behavior, including those situations which employ variously scheduled aversive stimuli. As in Estes' (1944) and Azrin's (1956) studies, the immediate effect of such an arrangement was a depression in the rate of response, i.e., number of words produced per 30-second interval. Furthermore, this effect differed depending upon the particular schedule employed, and was maintained as long as the negative reinforcer was administered. Discontinuation of the aversive stimulus resulted in almost complete recovery of response rate in two of the groups and in partial recovery by the third group. Thus in those cases where the elimination of verbal behavior is an objective, the present study suggests certain limitations in the employment of a secondary, aversive reinforcement.

#### SUMMARY

This study investigated the effect of a negative verbal cue upon verbal rate in a projective test-like situation. Sixty hospitalized patients were randomly assigned to one of four groups for the purpose of administering different schedules for the verbal cue: 20-second interval, every other 20-second interval, variable interval, and control. The stimulus materials were the TAT and the Symonds Picture Story Test. Each group

experienced three continuous phases. During the first 4 minutes verbal responses for the four groups were recorded and no differential treatment occurred. Following this, subjects in the experimental groups were exposed to 20 negative cues at the predetermined rate. When this criterion had been achieved, verbal responses were recorded for an additional 4 minutes with the cue withdrawn for all four groups.

It was found that the verbal cue employed in this study ("Unh unh") acted as a negative reinforcer. The influence of the different schedules were also revealed, with the most profound effect upon training and durability attributable to the variable interval schedule.

#### REFERENCES

- ADAMS, J. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, 54, 383-405.
- ADAMS, J. S., & HOFFMAN, B. The frequency of self-reference statements as a function of generalized reinforcement. *J. abnorm. soc Psychol.*, 1960, 60, 384-389.
- AZRIN, N. H. Some effects of two intermittent schedules of immediate and non-immediate punishment. *J. Psychol.*, 1956, 42, 3-21.
- DINSMOOR, J. A. Punishment: I. The avoidance hypothesis. *Psychol. Rev.*, 1954, 61, 34-46.
- EDWARDS, A. E. *Experimental design in psychological research*. New York: Rinehart, 1956.
- ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, 57(3, Whole No. 263).
- FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement*. New York: Appleton-Century-Crofts, 1957.
- GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, 68, 409-416.
- JENKINS, W. O., & STANLEY, J. C. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, 47, 193-234.
- KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, 35, 148-170.
- MURRAY, H. A. *Thematic Apperception Test manual*. New Haven: Harvard Univer. Press, 1943.
- SALZINGER, K. Experimental manipulations of verbal behavior: A review. *J. gen. Psychol.*, 1959, 61, 65-94.
- SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.*, 1940, 9, 371-377.
- SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.
- SYMONDS, P. M. *Adolescent fantasy*. New York: Columbia Univer. Press, 1949.

(Received March 21, 1961)



# YEASAYERS AND NAYSAYERS:

## A VALIDATING STUDY

IRWIN MAHLER

*Occidental College*

In a recent article, Couch and Keniston (1960) discussed the rationale and the construction of the Agreement Response Scale (ARS), designed to measure an agreeing response tendency which was hypothesized to be a manifestation of a significant underlying personality syndrome. Using an original Over-all Agreement Score (OAS) these investigators selected a group of 10 Yeasayers (high agreement scores) and a group of 11 Naysayers (low agreement scores) for clinical study. The clinical findings revealed that Yeasayers were individuals with weak ego control who seemed to have strong external orientations, being primarily responsive to group values and demands. Naysayers, on the other hand, were more internally oriented, more introverted, and exhibited greater capacity to inhibit and suppress impulses.

The present investigation may be considered an attempt to establish the construct validity of the Agreement Response Scale. If a group of individuals could be found with strong group orientation, including acceptance of group values and demands, it might be expected that they would score higher on the ARS than would individuals lacking this group orientation. It seemed reasonable that members of college sororities and fraternities could be called group oriented, and thus could constitute one known group. However, because such group loyalty could be a consequence of membership rather than a cause of it, it appeared more desirable to conduct this study with subjects not yet members of these groups. Two hypotheses which might easily be tested were formulated:

1. College students who indicate a desire to join fraternities and sororities (Rushees) will have higher ARS scores than will those students who do not desire to join these organizations (Nonrushees).

2. Of those who rush, those who receive bids (Pledges), that is, those who are offered membership in the fraternities and sororities, will have higher ARS scores than those who do not receive bids (Rushees-No Bid).

### METHOD

The subjects for this study were 219 men and 163 women, all Occidental College freshmen. During

the last week prior to the offering of bids by the fraternities and sororities, the ARS scale was administered in the various discussion groups of a required course for freshmen. The total of 382 freshmen who completed the scale represented almost the complete population of freshmen (395) in the college. Through the cooperation of the offices of the Deans of Men and Women it was possible to obtain the names of those freshmen who had signed up for, and began, the rushing procedure. There were 72 freshmen women and 104 freshmen men who could be designated Rushees, leaving 91 women and 115 men to be designated Non-rushees. The following week it was possible to obtain the names of the 30 freshmen women and 71 men who received bids; they were designated as the Pledges. The remaining 33 men and 42 women were designated as Rushees-No Bid. The apparent differences between men and women in rushing and pledging have been typically true in this college, reflecting the fact that the fraternities are national organizations but the sororities are local. It also is due to limitation of the size of sororities; that is, these groups are limited to a small number of pledges each year, whereas the fraternities are not so limited.

### RESULTS

The mean ARS scores for the 382 subjects were computed to be 55.24, with the *SD* being 9.7. For men, *N* = 219, the mean was 55.82 and the *SD* was 9.48; for women, *N* = 163, the mean was 54.45, the *SD* was 9.96. This difference between means is not significant (*CR* = 1.36), nor are any of the apparent differences between sexes presented for each subgroup in the table below.

Table 1 presents the data relevant to the first hypothesis. It can be seen that male rushees have significantly higher ARS scores than do non-

TABLE 1  
MEAN ARS SCORES OF RUSHEE  
AND NONRUSHEE GROUPS

	Rushees			Nonrushees			Difference	<i>t</i>
	<i>M</i>	<i>N</i>	<i>SD</i>	<i>M</i>	<i>N</i>	<i>SD</i>		
Men	57.64	104	9.55	54.17	115	9.10	3.47	2.73**
Women	55.38	72	9.60	53.72	91	10.30	1.60	1.07*

\* *p* > .10.

\*\* *p* < .01.



TABLE 2  
MEAN ARS SCORES OF PLEDGES AND  
RUSHEE-NO BID GROUPS

	Pledges			Rushlee-No Bid			Dif- ference	t
	M	N	SD	M	N	SD		
Men	59.01	71	9.95	54.70	33	7.64	4.41	2.40**
Women	57.27	30	8.95	54.02	42	9.47	3.25	1.47*

\*  $p > .10$ .

\*\*  $p < .02$ .

rushlee males, but this difference does not hold for women. Similarly, Table 2 shows that of those men who rush, those who receive bids have significantly higher ARS scores than those who do not pledge. Once again, this difference is not true for women.

It will be noticed that the men who do not receive bids have a mean ARS score (54.70) very similar to that of the men who do not rush (54.19), which would appear to further support the implicit hypothesis that fraternity men are different from nonfraternity men. The same relative similarity is true for women: those who do not receive bids have a mean score (54.02) close to that of those women who do not rush (53.72). Table 3 shows that when the differences between pledges and all those who do not pledge are tested, it is found that the difference is significant for men, but for women it just misses significance at the .05 level.

It appears that the predicted results hold true consistently for men, but not for women, even though there are no significant differences in ARS scores between the sexes. Inasmuch as Couch and Keniston (1960) originally standardized the ARS scale on men only, it may be that the scale contains a few items which are more differentiating for men than for women. The fact that the scale may be less differentiating for women than for men should not detract from the general verifi-

TABLE 3  
MEAN ARS SCORES OF PLEDGES AND  
ALL NONPLEDGES

	Pledges			All Nonpledges			Dif- ference	t
	M	N	SD	M	N	SD		
Men	59.01	71	9.95	54.29	148	8.80	4.72	3.73**
Women	57.27	30	8.95	53.81	133	10.06	3.46	1.84*

\*  $p < .10$ .

\*\*  $p < .01$ .

cation of the hypotheses about the differences between externally and internally oriented people. Further study of the applicability to women of the concept of an agreeing response tendency as an underlying personality syndrome seems to be in order.<sup>1</sup>

#### SUMMARY

Couch and Keniston (1960) described a personality syndrome which seemed to fit members of fraternities and sororities, as opposed to college men and women who do not join such groups. Predictions were made as to the ARS scores of those who will pledge these groups, those who rush but are not pledged, and those who do not rush. These predictions were supported for men but not for women.

<sup>1</sup> In a separate study done subsequent to this one, Lucy Beebe found a mean ARS score for 79 non-sorority women of 53.15, and a mean score of 57.38 for 89 sorority women, which produced a significant difference at the .05 level. The subjects in this study contained no freshmen, so there was no overlap between the subjects in the two studies.

#### REFERENCE

- COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *J. abnorm. soc. Psychol.*, 1960, 60, 151-174.

(Received April 4, 1961)

#### ERRATUM

On page 637 of the article "Pancultural Factor Analysis of Reported Socialization Practices," by Leigh Minturn Triandis and William W. Lambert (*J. abnorm. soc. Psychol.*, 1961, 62, 631-639), the authors found that the scoring of sex of child variable is such that this factor actually reflects mothers who permit their girls to be aggressive. Thus the word boy should be replaced by girl in the description of Factor 9. The correct wording is:

(Scale 2, +.48, reflecting the "degree to which mother is positive when child fights other children"), tied somewhat to girls . . . which gives a picture of an expressive, positive mother, who permits her girl to be expressive as well.

# Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

## CONSULTING EDITORS

JOSEPH B. ADELSON  
*University of Michigan*

GORDON W. ALLPORT  
*Harvard University*

NORMAN H. ANDERSON  
*University of California, Los Angeles*

LEONARD BERKOWITZ  
*University of Wisconsin*

JACK W. BREHM  
*Duke University*

URIE BRONFENBRENNER  
*Cornell University*

DONALD R. BROWN  
*Bryn Mawr College*

DONALD T. CAMPBELL  
*Northwestern University*

LAUNOR F. CARTER  
*System Development Corporation*

DWIGHT W. CHAPMAN, JR.  
*Vassar College*

MORTON DEUTSCH  
*Bell Telephone Laboratories*

SEYMOUR FESHBACH  
*University of Pennsylvania*

DONALD W. FISKE  
*University of Chicago*

NATHANIEL L. GAGE  
*University of Illinois*

NICHOLAS HOBBS  
*Peabody College for Teachers*

MURRAY HORWITZ  
*New York University*

IRVING L. JANIS  
*Yale University*

HAROLD H. KELLEY  
*University of California, Los Angeles*

HERBERT C. KELMAN  
*Harvard University*

DANIEL J. LEVINSON  
*Harvard Medical School*

DAVID C. MCCLELLAND  
*Harvard University*

ROBERT B. MACLEOD  
*Cornell University*

HAROLD L. RAUSH  
*National Institute of Mental Health*

NEVITT SANFORD  
*Stanford University*

IRVING SARNOFF  
*Western Reserve University*

WILLIAM A. SCOTT  
*University of Colorado*

JOSEPH VEROFF  
*University of Michigan*

JUSTIN L. WEISS  
*Massachusetts Mental Health Center*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is \$10.00, or \$20.00 annually (foreign, \$20.50). Single issues are \$2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ  
University of Michigan  
Box 1248  
Ann Arbor, Michigan

---

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR  
Acting Production Manager

JUDY SMITH  
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 Sixteenth Street, N.W.  
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
Prince and Lemon Sts., Lancaster, Pa.  
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Lancaster, Pa., and at additional mailing offices

Copyright © 1962 by the American Psychological Association, Inc.



## THE EFFECTS OF TRAINING NEGROES UPON COOPERATIVE PROBLEM SOLVING IN BIRACIAL TEAMS<sup>1</sup>

IRWIN KATZ AND MELVIN COHEN

*New York University*

In two recent experiments on biracial work groups it was found that Negroes behaved compliantly toward white partners. Katz, Goldston, and Benjamin (1958), and Katz and Benjamin (1960) assigned cognitive and motor tasks to groups composed of two Negro and two white Northern male college students. In both studies Negroes made fewer suggestions than whites, readily accepted the latter's proposals, and tended to ignore one another. These features of Negro behavior occurred even when the men were told that they would receive extra pay if they worked well together as a team, and that they all had higher ability than men in other teams. These features also occurred in the second experiment, despite careful matching of Negro and white teammates on intelligence, and the use of cognitive tasks on which their performances were made to appear objectively similar.

The present report describes an experimental attempt to modify Negro behavior toward white teammates in the direction of greater assertiveness and autonomy. The arrangements for affecting this change were derived from assumptions about the sources of Negro inhibition in biracial work groups. An important inhibitory factor, it would appear, is the Negro's sense of intellectual inadequacy in such situations. Katz and Benjamin (1960) found that Negroes who actually did as well as white teammates on mental tasks perceived their own performance as inferior. If Negroes see themselves as incompetent in mixed groups they may expect to be judged similarly by white partners. Hence, when disagreements occur they should tend to assume that their own opinions are probably wrong, and even if not wrong, likely to be rejected by the whites.

The Negro's participation might be further inhibited by fear of arousing hostility in white work mates. He might expect them to interpret his behavior as an aggressive denial of his traditionally low status, with its implications of low intelligence and deference to whites.

What conditions are likely to reduce these restraints on Negro behavior? The two previously mentioned experiments employed factors that are known to promote cohesiveness in all-white groups (group reward and high group prestige). The experimenters even used rigged tasks so that Negroes and whites would appear to have equal cognitive ability. But there were relatively minor effects on the level and quality of Negro participation in team discussions. Apparently, the Negro subject's general avoidance of assertive acts was so complete that he had no opportunity to discover that these acts may not, in fact, have had punitive consequences. On theoretical grounds one would expect that acquired avoidance responses can be weakened by forcing the opposite behavior in the absence of the original stimulus that produced anxiety (Dollard & Miller, 1950).

The present experiment applied this approach to Negro behavior in teams composed of one Negro subject and one white subject. Partners were required to agree on answers to a series of cognitive problems on which they had been matched for ability. After every team decision the experimenter announced the correct answer. By varying the information available to partners about most of the problems the experimenter created disagreements in which one man confidently proposed a correct solution, and the other suggested with low confidence a wrong solution. Each man was favored with better information than his partner on 50% of the problems. In teams put through this experience (Assertion Training condition) Negroes and

<sup>1</sup> This research was carried out under Contract Nonr-285(24) between the Office of Naval Research and New York University. The project is under the direction of Irwin Katz.

whites were expected to have equal influence on the determination of team answers.

In a No Training condition biracial dyads were subjected to the same procedure on the cognitive problems as described above, except that the amount of information given to the favored subject on each problem was not so great as to force him to propose the correct answer with a high degree of expressed confidence. It was expected on the basis of Katz and Benjamin's (1960) findings that Negroes, as compared with whites, would express less confidence in their proposed solutions, and would have less influence on team decisions.

Before and after working on the cognitive problems, subjects made judgments about characteristics of briefly presented visual stimuli. For every item they individually wrote down two guesses, the first of which was revealed to the partner before the second was made. Then subjects had to agree on a team guess, after which the experimenter stated the correct answer. The subject's influence was defined in terms of the similarity of team decisions to his first guesses. It was predicted that Negro influence on team judgments would (a) increase from the pretest to the posttest, under Assertion Training; and (b) decrease under No Training.

As Negroes in the Assertion Training condition experienced intellectual competence and were able to influence the white partner, any anxiety they may have had about interacting with the partner should diminish. Hence, their behavior should become more task oriented and less defensive. On the other hand, Negroes in No Training, whose experience would be opposite to that of Assertion Training subjects, should tend to become more defensive. One evidence of nondefensive task involvement on the part of Negro subjects would be an ability to utilize the informational value of the partners' judgments, i.e., to accept them when their accuracy was probably high and to ignore them when their accuracy was probably low. This ability would be seen if the subject's second guess on any item showed movement toward the partner's first guess in proportion to the partner's known accuracy on previous trials. Accordingly, we predicted that the relation-

ship between the Negro's second guess movement toward the partner, and the difference between his own and his partner's accuracy on previous items, would (c) under Assertion Training increase from the pretest to the posttest, and (d) under No Training decrease from the pretest to the posttest.

Although our interest was focused mainly on Negro behavior we used white subjects, instead of holding white behavior constant by employing a confederate. This was done because it was not known, specifically, how whites would react to Negroes in the experimental situations. We could only make the general assumption that their behavior would not differ so much under Assertion Training and No Training as to cancel out the anticipated effects on Negro behavior. A further reason for using white subjects was to permit exploration of white reactions to Negroes under these conditions.

#### PROCEDURE

*Subjects.* All subjects were male students in New York City colleges and universities. Fifty Negro volunteers and 65 white volunteers were tested individually on a set of mental problems. Then the 36 highest scoring Negroes were paired with 36 whites on the basis of score similarity, and the pairs were randomly assigned to the two experimental conditions. The age range of Negro subjects was 18-35 and the mean age was 22.3. All except two Negroes had lived in the North for at least 3 years. In the white sample the age range was 18-30 and the mean age was 21.

*Cognitive problems.* There were 49 cognitive problems mounted on 3" x 5" cards. Twenty-five of them were adapted from Raven's Standard Progressive Matrices and the remainder were adapted from various other nonverbal mental tests. The problems were all of the fourfold multiple-choice type. Negroes and whites had different sets of problem cards. On 20 problems the Negro version was easy and the white version was unsolvable, and on 20 problems this was reversed. There were identical versions of 9 problems, of which 6 were easy and 3 were unsolvable.

*Procedure for cognitive problems.* Each subject worked on the same set of cards twice, first at an Individual session and about 2 weeks later at a Team session. At the Individual session the problems were introduced with the explanation that they were being tried out for possible use in an aptitude test. The subject was offered a bonus of \$.35, in addition to the regular pay of \$1.50, for achieving a high score. The scoring system required the subject to "bet" zero to five points on each answer, depending upon his degree of confidence. He was told that the pur-



pose of the betting was to provide information about the relationship between confidence and correctness.

Time limits of 15 seconds for easy items and 6 seconds for unsolvable items insured that most of the easy items were answered correctly and with confidence, and that the unsolvable items aroused confusion. Subjects were not told how many items they had answered correctly, but merely that they had done well. All who were to be recalled were awarded the bonus.

Negroes and whites who had made about the same number of errors on easy items were paired in Team sessions. Partners were told that they had equal cognitive ability. A bonus of \$35 was offered to each man "if the two of you do better as a team than you did as individuals." Subjects were given the same cards that they had worked on previously, and informed that the two sets were identical. When the experimenter signaled, they turned up their corresponding cards for each problem, and viewed it for 6 seconds. Then they had to arrive at a single team answer and bet through free discussion. Finally, the experimenter announced the correct answer.

Under Assertion Training, the men had sheets of paper which contained their "previous answers and bets." Their actual responses had been altered so that correct answers with high bets (4 or 5 points) were made to all easy items and incorrect answers with low bets (0 or 1) were made to all unsolvable items. Before looking at a problem, each subject was to check his previous response to that item, so that he could then confirm his original response or revise it. After seeing the problem, the men were required to start their discussion by reading aloud their previous answers and bets, which they were then free to abandon.

Under No Training, subjects were not given information about their previous responses to the problems. Instead, both men had blank sheets on which they wrote individual answers and bets to each problem. Team discussions began with partners announcing to one another what they had just written.

*Judgmental task.* At the Team session, both before and after they worked on the cognitive problems, the men were shown comparable sets of 11 illustrations, each consisting of 4 photographs of people, cows, balloons, etc.; 4 drawings of irregularly shaped lines; and 3 drawings of enclosed figures. Subjects had to estimate the number of objects, the length of lines, and the area of figures. Each illustration was briefly exposed to both men. Then they wrote down their first guesses, which the experimenter read aloud. The stimulus was shown briefly again, and the subjects wrote their second guesses. These were not read aloud. There was a third brief presentation of the stimulus, after which the subjects had to arrive at a consensus. Finally, the experimenter gave the correct answer.

The subject's influence was scored as the number of times the team decision was closer to his own first guess than to the partner's first guess; ties were

scored as one-half.<sup>2</sup> Movement of the subject's second guess toward the partner's first guess was expressed as a percentage of the total distance between the subject's first guess and the partner's first guess. No movement, or movement away from the partner, was scored as zero, and movement up to or beyond him was scored as 100.

Accuracy of individual guesses was measured by means of decile scores, which for each item were based upon a single distribution of the raw error scores (i.e., discrepancies between first guesses and the true item value) obtained by all subjects in the experiment on that item.

To evaluate relationships between second guess movement and partners' previous accuracy we computed Pearson  $r$ 's for the subject's movement on Item 4 (the last of the "number of objects" type) vs. the total difference between own and partner's accuracy on Items 1-3. Items 5-11 were not used because of marked within-subject variability of the accuracy scores on the three types of items (aggregates, lines, and areas). The variability attenuated relationships between movement on later items and previous accuracy.

*Terminal questionnaire.* This contained items on perception of own accuracy and partner's accuracy on the two tasks, perception of influence, and preference for the same partner in a future experiment.

## RESULTS

### *Adequacy of the experimental arrangements.*

Negroes and whites under Assertion Training were supposed to have equal control over team solutions to cognitive problems. Under No Training, Negroes were supposed to be less confident, and to control fewer team decisions, than whites. Means and  $t$  values that reveal whether these experimental specifications were met are presented in Rows 1 and 2 of Table 1. Row 1 shows that Negroes under No Training did express significantly less confidence (i.e., bet fewer points) in their proposed solutions to easy problems than did whites. (In Assertion Training the mean size of bets was controlled at 4.7.)

Row 2 in Table 1 refers to the number of problems on which initial disagreement was resolved in favor of Negro or white proposals. Because scores were available from both ex-

<sup>2</sup> A more exact scoring method was also tried, in which the Negro subject's movement on each item was expressed as a percentage of the amount of movement by both men combined, and his percentage scores on all 11 items were summed. This type of score had very low odd-even reliability, whereas the dichotomous scoring method that we adopted had a median odd-even  $r$  (uncorrected) of .88.



TABLE 1  
MEAN SCORES ON INFLUENCE, CORRECTNESS, AND SIZE OF BETS ON COGNITIVE PROBLEMS <sup>a</sup>

Type of mean score	Assertion Training			No Training		
	Whites	Negroes	<i>p</i>	Whites	Negroes	<i>p</i>
Amount the subject proposed betting on own easy items	4.70	4.70	—	4.41	3.81	<.001
Number of times own proposal accepted by partner	20.56	19.89	>.20	17.17	14.00	<.01
Change in number of errors from Individual session to Group session	—	—	—	-.67	1.00	<.05

Note.—*N* = 18 for all means.  
<sup>a</sup> The *p* values are based on *t* tests, using both tails of the distribution; on the first and third rows, Assertion Training scores were held constant for Negro and White subjects, and *t*'s for No Training subjects were based on *SE<sub>MD</sub>*; on the second row, *t* tests were derived from an analysis of variance.

perimental conditions they were first submitted to an overall statistical test, a 2 × 2 analysis of variance (not referred to in Table 1). This analysis revealed a significant (*p* < .001) Race × Condition interaction, which justified the use of *t* tests to evaluate the differences between races under each condition. As the means and *t* values in Row 2 of the table indicate, partners in Assertion Training did not differ in the amount of influence they had over team decisions, but in No Training, whites had significantly more influence than Negroes. Thus, it is apparent that the manipulation of Negroes' successful assertion was adequately accomplished.

Table 1 also shows that *both* races had more influence on team decisions in Assertion Training than in No Training. This difference between conditions was a significant main effect (*p* < .001) in the analysis of variance mentioned above, and can be attributed to the fact that subjects were not as closely guided on easy items in No Training as they were in the other condition. That is, unlike subjects in Assertion Training, these men were not actually given correct answers and high bets on easy items, and then required to read these aloud to the partner. Hence, the person with the easy version of a problem was not always sufficiently confident of his answer to win the decision.

Row 3 in Table 1 refers to differences in the number of individual errors made on easy problems by No Training subjects in the Individual and Team sessions. (In Assertion Training, subject's answers in the Team session were experimentally controlled.) Negro

errors increased in the Team session, while white errors decreased, and this race difference was significant (*p* < .05). Such evidence of cognitive disruption on the part of No Training Negroes is consistent with results in Rows 1 and 2, and further suggests that the presence of a white partner in the Team session was stressful for them.<sup>3</sup>

Also bearing on the question of experimental adequacy was the constancy of individual accuracy in the judgmental task. Analysis of variance of pre-post differences in accuracy scores revealed that Race, Conditions, or Race × Conditions did not consistently affect accuracy on the three types of judgmental items. Hence, there was no possibility of the dependent variables being confounded with accuracy.

*Tests of the predictions.* It was predicted that Negro influence on team judgments would (a) increase under Assertion Training, and (b) decrease under No Training. Table 2 contains the mean influence scores of Negroes, and results of *t* tests. (Scores of whites are obtainable by subtracting Negro scores from 11.) Pretest scores of Negroes in the two experimental conditions were not significantly different, and the combined pretest means were slightly below 5.50, the point at which

<sup>3</sup> The possibility exists that the increase in Negro errors was a regression phenomenon, since the 36 Negro subjects were selected as the highest scorers out of 50 men in the Individual session. However, inasmuch as the selection included more than two-thirds of the original pool this possibility would appear to be small. Moreover, comparison of the subjects' error scores in the two sessions revealed no relationship between original scores and amount of increase in errors.

Negroes and whites had equal influence.<sup>4</sup> Then, in the posttest, the Assertion Training mean rose significantly ( $p < .05$ , one-tailed) and the No Training mean declined significantly ( $p < .05$ , one-tailed).<sup>5</sup> Thus, Predictions *a* and *b* were confirmed.

It was predicted that Negro subjects' second guess movement toward the partner would (*c*) under Assertion Training be increasingly related to own and partner's previous accuracy, and (*d*) under No Training be decreasingly related to it. As described under Procedure, the subjects' second guess movement on Item 4 was correlated with partner's accuracy-minus-own-accuracy on Items 1-3. The obtained *r*'s for whites and Negroes, pretest and posttest, are presented in Table 3. Only one correlation was significant—that for Assertion Training Negroes in the posttest ( $r = .87$ ,  $p < .001$ ). It indicates that these Negro subjects moved toward their white partners on Item 4 to the extent that the latter's accuracy exceeded own accuracy on previous items. The correlation was significantly higher ( $p < .01$ ) than the same subjects' pretest correlation, in confirmation of Prediction *c*.<sup>6</sup> Because the pretest correlation

<sup>4</sup> The fact that the combined pretest means of Negroes were only slightly below the point of equal influence might appear to contradict the previous findings of Negroes having less influence on biracial decisions. We believe that the present task of guessing characteristics of briefly exposed stimuli was not initially seen by the subjects as related to intelligence or any other type of socially valued capacity, while in previous studies the tasks were always either cognitive problems or human relations problems.

<sup>5</sup> One-tailed tests were used to test predictions. All other tests used both tails of the distribution.

<sup>6</sup> Relevant to these findings is the question of changes from the pretest to the posttest in the total

TABLE 2

MEAN INFLUENCE SCORES OF NEGROES ON JUDGMENTAL TASK BEFORE AND AFTER WORKING ON COGNITIVE PROBLEMS

Condition	Pre <sup>a</sup>	Post	Difference
Assertion Training	5.00	5.93	.93**
No Training	5.50	4.88	-.62**

Note.— $N = 18$  for all means.

<sup>a</sup> For the "pre" difference between Assertion Training and No Training means,  $p > .20$ , based on a *t* test.

\*\*  $p < .05$ . The  $p$  values are based on *t* tests of the mean differences, using one tail of the distribution.

TABLE 3

CORRELATIONS (*r*) BETWEEN SECOND GUESS MOVEMENT TOWARD PARTNER ON ITEM 4 AND PARTNER'S ACCURACY MINUS OWN ACCURACY ON ITEMS 1-3

	Assertion Training			No Training		
	Pre	Post	<i>p</i>	Pre	Post	<i>p</i>
Negro	.05	.87**	<.01 <sup>a</sup>	-.02	-.17	<i>ns</i>
White	.23	.28	<i>ns</i>	.39*	.09	<i>ns</i>

Note.— $N = 18$  for all correlations.

<sup>a</sup> Based on *t* test of *z* scores; value of *p* the same for one or both tails.

\*  $p < .20$  (both tails).

\*\*  $p < .001$  (both tails).

for No Training Negroes was close to zero, Prediction *d* was not testable.

Only one *r* for white subjects was even suggestive of a relationship—that for No Training subjects in the pretest ( $p < .20$ ).

*Perceptions of performance.* Differences in the subjects' perceptions of self and partner were evaluated by means of the Wilcoxon matched-pairs test. (A nonparametric test was used because of heterogeneous score distributions.) Table 4 indicates that the experimental conditions did not affect whites' relative perceptions of self and partner respecting (*a*) accuracy on mental problems, (*b*) influence on mental problems, and (*c*) influence on the judgmental task. Under both conditions they saw themselves as answering more problems correctly than did their partners, but having only equal influence. Actually, whites' accuracy was higher only in No Training, and Negro influence was equal only in Assertion Training. Negroes, on the other hand, saw (correctly) under No Training that own accuracy (*a*) and own influence (*b*) on the problems were less than the partner's, and under Assertion Training they perceived (correctly) that self and partner were equal both in accuracy and influence.

On Topic *c* of Table 4, the failure of the subjects' answers to reflect actual differences in influence on the judgmental task under the

amount of second guess movement by the subjects, since the occurrence of such changes could necessitate a qualification of the interpretation. The mean movement scores on Item 4 were found by analysis of variance to be similar for all combinations of race, experimental conditions, and pre- or posttest.

TABLE 4  
SUBJECTS' MEAN ESTIMATES OF OWN AND PARTNER'S PERFORMANCE

Items and $p$ levels <sup>a</sup>	Whites		Negroes	
	Assertion Training	No Training	Assertion Training	No Training
A. Accuracy on Problems				
1. On 49 items, guess number you answered correctly.	32.5	29.0	31.3	27.6
2. Same for partner.	30.5	26.2	31.5	31.0
$p$	<.05	<.02	<i>ns</i>	<.01
B. Influence on Problems				
3. How often accepted partner's answer?	15.4	16.6	19.6	16.0
4. How often he accepted yours?	16.3	12.9	19.7	12.1
$p$	<i>ns</i>	<i>ns</i>	<i>ns</i>	<.01
C. Influence on Judgmental Task				
5. On 22 team decisions, how often mainly followed partner?	9.0	9.4	10.8	9.8
6. How often he mainly followed you?	9.1	10.3	9.9	8.8
$p$	<i>ns</i>	<i>ns</i>	<.05	<.05

<sup>a</sup> All  $p$  values (both tails) are for differences between responses to paired items using the Wilcoxon matched-pairs test,  $N = 18$ .

two experimental conditions may simply reflect a defect in Items 5 and 6, which referred to all judgmental trials, half of which occurred before the experimental inductions. A questionnaire item, not shown in Table 4, asked "Which of you made more accurate judgments?" Evaluation of responses by means of the binomial test resulted in only one significant finding—Negroes in No Training attributed (incorrectly) greater accuracy to the partner ( $p < .05$ ).

*Acceptance of partner.* A question dealt with preference for the same partner if called back to work again. Dichotomized replies were subjected to chi square tests of association with the two experimental conditions. Negroes, it was found, expressed the same degree of preference for the partner in both conditions. But whites in Assertion Training preferred their partners less than did whites in No Training ( $p < .01$ ). Sign tests were used to compare responses of teammates, and revealed that whites in Assertion Training tended to prefer their Negro partners less than their Negro partners preferred them ( $p < .10$ ).

#### DISCUSSION

In this experiment the rise in Negro dominance in Assertion Training contrasts sharply with the loss of influence of Negroes in No Training. Yet in both conditions partners were given alternately easy and hard versions

of problems, to facilitate equal participation in team decisions. Only when forced, did Negroes display competence. But once having done so, they were able to maintain a high level of behavioral autonomy on another task. These findings suggest that Negroes experienced anxiety in the biracial situation, which was reduced to some extent by the behavioral training.

Findings on individual influence over team decisions in the judgmental task refer to outcomes of dyadic interaction; hence, they simultaneously reflect changes in the tendencies of both partners to yield or to dominate. It cannot be stated whether the rise in influence of Negroes after training was due mainly to greater assertiveness on their part, or to the development in whites of greater compliance. The latter possibility is suggested by Hochbaum's (1954) experiment. He raised or lowered the confidence of the subjects by telling them they had done well or poorly on a test; those who had their confidence reduced, yielded more quickly to subsequent group pressures on related matters than those whose faith in themselves had been raised. Our white subjects in Assertion Training may have experienced a lowering of self-esteem when they saw their own performance equaled by a member of a social group which has been stereotyped as of low intelligence.

The findings on second guess movement



show that the behavior of Negroes did change after training, and that they now seemed to respond privately to partners' guesses in accordance with their probable accuracy. Thus, Negroes after training apparently reacted to their partners' opinions with objectivity. It is interesting to speculate that, in general, when Negroes acquire job roles with decision making responsibilities they may become highly concerned about giving "fair" consideration to the views of white workers, as a means of maintaining satisfactory relations with them. The lack of correlation between second guess movement and previous accuracy of partner that was found for No Training Negroes, and for all whites, would indicate that these subjects were responding mainly to other characteristics of the partner, such as his racial identity.

The data strongly suggest that the training induction aroused defensiveness in whites. On the questionnaire, whites in Assertion Training downgraded the problem solving performance of their Negro partners, yet saw them as having equal influence on team solutions. Also, these subjects were less willing than No Training whites to work again with the same partner. They may have felt threatened in two ways: first, as suggested above, the similarity between their own and their Negro partner's performance may have wounded their pride; second, the rising influence of the partner may have been experienced as a deprivation of power. Presumably, both of these effects would depend upon the content and salience of the white's stereotype of Negroes.

It is noteworthy that Negroes in Assertion Training correctly saw their own accuracy on cognitive problems as equal to their partners, while Katz and Benjamin (1960) found that when whites and Negroes were made to perform equally well on cognitive tasks, Negroes viewed their own performance as inferior to that of their companions. Although the two studies are not strictly comparable, the critical factor underlying this difference apparently is that in the previous experiment Negroes were not forced to participate actively in team decisions, and tended to yield to

white pressure even when their own judgments were correct.

#### SUMMARY

Biracial dyads, composed of Negro and white Northern college students, engaged in cooperative problem solving. Unknown to them, subjects received different information, so that only one could solve any problem. Each subject had the soluble version half the time. Correct solutions were announced after each team decision. Under Assertion Training the partner with the easy version could propose the correct answer with high expressed confidence. Under No Training the person having the easy version was not forced to propose the correct solution. Pre- and postmeasures of social influence were obtained on another task.

It was predicted (a) that Negro influence would increase in Assertion Training and (b) decrease in No Training, and (c) that the relation between movement of Negroes' private judgments toward the partner's judgments, on the one hand, and the partner's previous accuracy, would increase in Assertion Training and (d) decrease in No Training. The first three predictions were upheld by the data, while the last was not testable. Other findings were: Negroes tended to make accurate comparisons of their own and their partner's ability and influence; whites in Assertion Training downgraded the partner's ability, and tended to reject him as a future co-worker. The last two findings suggest that the confident show of ability by Negroes in Assertion Training was ego threatening to whites.

#### REFERENCES

- DOLLARD, J., & MILLER, N. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- HOCHBAUM, G. The relation between group members' self-confidence and their reactions to group pressures to uniformity. *Amer. sociol. Rev.*, 1954, 19, 678-687.
- KATZ, I., & BENJAMIN, L. Effects of white authoritarianism in biracial groups. *J. abnorm. soc. Psychol.*, 1960, 61, 448-456.
- KATZ, I., GOLDSTON, JUDITH, & BENJAMIN, L. Behavior and productivity in bi-racial work groups. *Hum. Relat.*, 1958, 11, 123-141.

(Received March 14, 1961)

## A COMPARISON OF THREE SETS OF ROTATED FACTOR ANALYTIC SOLUTIONS OF SELF-CONCEPT DATA

PHILIP A. SMITH<sup>1</sup>

*System Development Corporation, Santa Monica, California*

Factor analytic evidence for the multidimensionality of the self-concept was presented by the writer in earlier papers (Smith, 1959, 1960). Utilizing a series of bipolar adjectives as a self-rating device, and factoring a matrix of phi coefficients based on these data, a major self-esteem dimension and four other factors descriptive of the self-concept in psychiatric patients were isolated. The present paper reports the factoring and objective rotation of additional self-concept data obtained from independent, equated groups of subjects.

With the exception of some work reported by Osgood, Suci, and Tannenbaum (1957) in which the self was one of several concepts studied, no replicated factor analytic studies of the self-concept have been reported. In general, studies which employ the principle of replication are infrequent; for factor analysis in particular, replication is seldom attempted if for no other reason than the computational labor required. But with the development of correlation and factor analytic programs for digital computers, such work becomes in practice fairly simple.

### METHOD

In the present study, two equated groups of subjects independently made self-ratings; data for the two groups were separately factor analyzed, then each factor matrix was rotated by three objective analytic procedures recently developed (Carroll, 1958; Kaiser, 1958; Neuhaus & Wrigley, 1954). The aim was to find rotational solutions which were essentially the same for both groups, and thus provide "cross-validated" dimensions of the self-concept. The results also provide indirect evidence

of the relative utility of each of the three rotational procedures used.

**Subjects.** Two samples of 96 male hospital patients equated for age, educational level, and gross hospital status, made self-ratings on 40 bipolar adjective scales (the semantic differential). Seventy-seven of the subjects in each sample were psychiatric patients, tested on admission or just prior to discharge from the psychiatric ward of a Veterans Administration general medical and surgical hospital. Each group also included 8 nonpsychiatric patients hospitalized for various medical problems, and 11 nonhospitalized persons (former psychiatric patients), who visited the hospital on an outpatient basis. A more complete description of the testing procedure and the nature of the patient population from which these subject samples were drawn is available in the reports of other studies which took place concurrently (Bostian, Smith, Lasky, Hover, & Ging, 1959; Goldberg & Smith, 1958; Lasky, Hover, Smith, Bostian, Duffendack, & Nord, 1959; Smith, Barrows, & Whitney, 1959).

**Procedure.** The ratings on all 40 bipolar scales were intercorrelated by the Pearson product-moment method and factor analyzed for each group separately. The classical centroid method was used, with communalities estimated as the highest correlation in each row and replaced by the highest residual for successive factors. Ten factors were extracted, and the first eight were selected for rotation.<sup>2</sup> Each factor matrix was rotated by the quartimax method (Neuhaus & Wrigley, 1954), the normal varimax method (Kaiser, 1958), and the biquartimin method (Carroll, 1957, 1958), all of which have been adapted for the computer. The first two procedures provide orthogonal estimates of simple structure, while the biquartimin (or oblimin) method gives an oblique solution. All factors in each pair of rotated solutions were compared by an objective procedure which yields an index of similarity for each set of factors compared. This measure has been variously termed *unadjusted correlation* (Barlow & Burt, 1954), *coefficient of congruence* (Tucker, 1951), and *degree of factorial similarity* (Wrigley & Neuhaus, 1955).<sup>3</sup>

<sup>1</sup> The author is indebted to E. Lowell Kelly for many helpful suggestions during this study. He also gratefully acknowledges the cooperation of the staff at the Ann Arbor Veterans Administration Hospital, and the assistance provided by the computing laboratories at the University of Michigan and the System Development Corporation.

<sup>2</sup> Only the first seven factors were rotated by the biquartimin method, since the eighth factor accounted for so little variance and appeared to be of negligible importance in determining orthogonal structure.

<sup>3</sup> Harman (1960, pp. 256-259) offers a more extended discussion of this and other factor matching procedures.



## RESULTS

*Equation of the groups.* Table 1 shows the means and standard deviations of the self-ratings. In no instance was there a significant difference between the two groups. The correlation matrices and unrotated factor structure based on these data (not shown) also failed to show appreciable differences. Possibly contributing to the close similarity of the groups was the tendency of subjects to rate themselves toward the socially desirable end of most continua. In Table 1 note, for example, the tendency for the self to be rated "kind," rather than "cruel" or neutral, and "true" rather than "false." Ambiguous or equivocal terms such as "cold-hot," having less obvious social implications, evoked ratings which clustered more closely about the objective midpoint of the scale.

*Quartimax solutions.* There are several obvious and troubling dissimilarities in these results. Factor I accounts for 28% of the variance in the first group. Inspection of the items with high loadings show this factor to be a general evaluative dimension. But, in the second group, these items load heavily on two evaluative factors (I and II), which account for 11 and 20% of the variance, respectively.<sup>4</sup> This finding is more readily apparent in Table 2, which summarizes the relationships between all factors for the two groups. The indices of similarity in this table may be interpreted as measures of correspondence or unadjusted correlation between pairs of factors from the two samples. Figures which are presented in bold face type indicate a statistically significant degree of correspondence between pairs of factors (see Kumata & Schramm, 1956). Factors I and II for the second group "correlated" highly with Factor I for the first group. Inspection of the items does not suggest any clearcut difference in the psychological meaning of these various

<sup>4</sup> Detailed findings can be found in Tables A, B, C, and D which have been deposited with the American Documentation Institute. Order Document No. 6941 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1  
MEAN SCORES OF THE TWO GROUPS  
ON ALL RATING SCALES

Variable	M <sup>a</sup>		SD	
	Group 1	Group 2	Group 1	Group 2
1. Shaky-Steady	2.80	2.90	1.18	1.13
2. Strong-Weak	2.80	2.56	1.03	.93
3. Rugged-Delicate	2.85	2.74	.83	.85
4. Sick-Healthy	3.34	3.27	1.13	1.09
5. Hard-Soft	3.03	3.01	.84	.91
6. Heavy-Light	3.22	3.00	.99	.96
7. Bad-Good	3.70	3.70	.89	.96
8. Tired-Refreshed	2.84	2.73	1.07	1.09
9. Big-Little	2.88	2.94	.92	.84
10. Nice-Awful	2.45	2.46	.94	.98
11. Masculine-Feminine	2.00	1.82	.95	.93
12. Underweight-Overweight	2.80	2.97	1.02	1.10
13. Empty-Full	3.15	3.41	1.06	.96
14. Kind-Cruel	1.88	1.89	.88	.90
15. Strange-Natural	3.62	3.52	1.22	1.12
16. Useful-Useless	2.26	2.07	1.16	1.00
17. Confident-Unsure	2.62	2.68	1.33	1.25
18. Success-Failure	2.73	2.72	1.31	1.19
19. Dangerous-Harmless	4.21	4.25	.93	.80
20. Happy-Sad	2.68	2.71	1.25	1.14
21. Worthless-Valuable	3.66	3.56	1.06	1.02
22. Stable-Unstable	2.63	2.73	1.23	1.14
23. Destructive-Constructive	3.92	3.96	.99	.89
24. Active-Passive	2.39	2.34	1.14	1.06
25. Afraid-Unafraid	3.37	3.41	1.25	1.20
26. Oversexed-Undersexed	2.98	3.14	.77	.68
27. Friendly-Unfriendly	1.91	2.14	1.07	1.04
28. Embarrassed-At ease	3.08	3.15	1.15	1.13
29. Loved-Hated	2.08	2.09	1.12	1.09
30. Satisfied-Dissatisfied	2.84	2.84	1.35	1.22
31. Angry-Calm	3.56	3.53	.94	1.09
32. Cold-Hot	3.05	3.17	.72	.78
33. Leader-Follower	2.98	3.02	1.01	1.02
34. Sharp-Dull	2.66	2.64	.91	.93
35. Nervous-Relaxed	2.35	2.42	1.09	1.11
36. Superior-Inferior	3.01	2.87	.84	.92
37. Smart-Dumb	2.52	2.46	.87	.92
38. Popular-Unpopular	2.55	2.62	1.04	.95
39. True-False	2.07	2.01	1.07	.92
40. Shy-Outgoing	2.82	2.79	1.06	1.04

<sup>a</sup> Each scale had five step intervals and was scored from 1 to 5. The objective midpoint for all adjective pairs is 3.0.

evaluative dimensions. Reference to Table 2 also reveals that only one dimension was clearly identified and replicated in both analyses. This was the factor describing body size (heavy, overweight, big versus light, underweight, little). Three other sets of factors appeared similar by inspection, but their computed indices of similarity were not sufficiently high to warrant them independently verified. It is important to note that lacking two sets of data, good arguments could have been marshaled for different but psycholog-



TABLE 2

INDICES OF FACTOR SIMILARITY: COMPARISON OF QUARTIMAX ROTATIONS IN TWO EQUATED GROUPS<sup>a</sup>

Group 1 rotated factors	Group 2 rotated factors							
	I Happy, Successful, Satisfied	II Useful, Loved, Good	III Heavy, Overweight, Big	IV Hard, Rugged, Strong	V Smart, Sharp, Leader	VI Nervous, Shaky, Tired	VII Oversexed, Nice	VIII Shy, Embarrassed, Inferior
Useful, I Successful, Loved	83	85	-24	38	52	-38	12	-22
Heavy, II Overweight, Big	-11	-02	87	-22	-07	25	06	08
Strong, III Hard, Rugged	30	-11	-25	53	15	-27	09	06
Smart, IV Sharp	-06	09	11	27	53	-07	06	-13
Angry, V Nervous	-30	-10	16	11	-14	49	10	17
Masculine, VI Nice	05	23	-01	-04	-10	07	46	26
Afraid, VII Cold, Sick	-25	-23	18	-11	-09	42	16	27
Leader, VIII Superior	15	-31	-10	11	39	-02	00	-20

<sup>a</sup> Decimals omitted.

ically meaningful descriptions of either solution.

*Normal varimax solutions.* Kaiser (1955, 1956) has criticized the quartimax method and offered a procedure which seems to have both theoretical and empirical advantages. In brief, his normal varimax method tends to maximize loadings by factors rather than by variables, and normalizes vectors during the analysis. He gives evidence that refactoring with additional variables may be accomplished without seriously influencing the size or distribution of loadings of an initial solution based on a fairly small number of variables.

Tables 3 and 4 show the normal varimax solutions for the two groups. It will be noted that there is a more even distribution of the variance contributed by each of the factors than was shown by the quartimax solutions. Reference to Table 5 shows that in other

respects also, the normal varimax solutions represent more satisfying resolutions of the data. The disappointing lack of congruence that was noted for the two quartimax solutions is not found. In Table 5, the indices of similarity for six pairs of factors are well above the lower limits of good matching, and may be considered cross-validated. Both groups show two major evaluative dimensions (Factors I and II), but the differences between these two factors can be somewhat more easily recognized. However, in both groups, these factors correlate rather highly with the remaining factors, suggesting that true orthogonality has not been attained.

*Biquartimin solutions.* While the approximation of simple structure attained by the varimax method allowed highly meaningful interpretations of the data, it seemed reasonable to expect that an oblique rotation of the data might also provide a meaningful solution.

TABLE 3  
NORMALIZED VARIMAX ROTATION FOR GROUP 1

Variable	Factors <sup>a</sup>								h <sup>2</sup>
	I	II	III	IV	V	VI	VII	VIII	
1. Shaky-Steady	00	-16	-21	-24	-08	29	-36	04	35
2. Strong-Weak	21	01	27	70	09	-29	00	00	70
3. Rugged-Delicate	15	10	19	54	11	-08	17	-11	42
4. Sick-Healthy	-05	-11	-09	-25	-06	62	-25	02	54
5. Hard-Soft	16	-12	-09	55	03	-18	-03	07	39
6. Heavy-Light	07	-06	88	07	09	-08	02	-07	81
7. Bad-Good	-13	-58	01	-17	05	16	-18	-20	48
8. Tired-Refreshed	-25	-05	-07	-24	-11	59	-30	15	60
9. Big-Little	09	20	42	17	40	09	14	01	44
10. Nice-Awful	27	61	10	-06	17	-12	13	-02	52
11. Masculine-Feminine	06	23	14	03	16	-25	10	-29	26
12. Underweight-Overweight	07	09	-83	-05	06	11	-13	-05	74
13. Empty-Full	-42	-04	-10	-11	-04	41	-12	-01	38
14. Kind-Cruel	06	67	-10	-06	12	-04	-05	-15	51
15. Strange-Natural	-51	-19	-05	-13	-11	24	-49	13	64
16. Useful-Useless	55	50	14	32	30	-02	09	03	77
17. Confident-Unsure	50	37	01	18	35	-20	-02	10	59
18. Success-Failure	60	28	04	29	34	-26	-05	14	73
19. Dangerous-Harmless	-24	-45	05	09	08	06	-24	-22	39
20. Happy-Sad	55	03	-16	35	17	-48	00	14	73
21. Worthless-Valuable	-55	-26	05	-15	-26	12	-20	-39	67
22. Stable-Unstable	57	27	-07	16	26	-13	25	-22	62
23. Destructive-Constructive	-12	-22	-05	07	-02	11	-67	-10	54
24. Active-Passive	53	12	05	10	37	-02	40	-03	61
25. Afraid-Unafraid	-12	07	03	-07	-18	64	-20	-14	53
26. Oversexed-Undersexed	01	-06	00	02	01	04	-04	-56	32
27. Friendly-Unfriendly	32	44	02	12	19	01	16	09	38
28. Embarrassed-At ease	-26	02	-16	-17	-19	37	-51	03	56
29. Loved-Hated	44	34	-12	-08	45	-17	36	-02	69
30. Satisfied-Dissatisfied	54	17	04	24	16	-47	04	-09	64
31. Angry-Calm	-18	-23	02	-09	-21	55	02	-32	54
32. Cold-Hot	05	20	-14	-10	-09	12	-33	-13	22
33. Leader-Follower	16	-14	17	06	80	-16	08	-04	75
34. Sharp-Dull	13	30	-09	46	66	-13	10	11	80
35. Nervous-Relaxed	-05	-08	-13	06	-23	58	00	-06	42
36. Superior-Inferior	26	04	-03	03	55	-38	01	-28	60
37. Smart-Dumb	02	41	-07	45	56	-13	05	15	73
38. Popular-Unpopular	29	22	04	-06	67	-28	14	-04	69
39. True-False	20	27	10	-05	40	-32	06	03	39
40. Shy-Outgoing	-23	15	02	-18	-49	23	04	-36	53
Percentage of variance	9.8%	7.8%	5.2%	5.9%	9.8%	8.9%	5.2%	2.6%	55%

<sup>a</sup> Decimals omitted.

Of the various rotational procedures discussed, Harman (1960) suggests that the biquartimin method is the most elegant mathematically. The factor structure following rotation by Carroll's biquartimin method shows that the size of the factor loadings are lower, on the average, than those generated by the other methods, but the differentiation of factors is somewhat clearer than the quartimax solutions.<sup>5</sup> Table 6 provides the indices of similarity between the factors for the two

groups. The similarities are not striking, and in most instances nonsignificant. Again it is suggested that had the results for either group alone been considered, somewhat different descriptions of factors would have resulted.

#### DISCUSSION

*Rotational methods.* Contemporary writers have taken opposed positions with regard to factor analysis. More serious than the criticism that one "gets out" only what he has "put in" is the argument that the results are

<sup>5</sup> See Footnote 4.

to a considerable degree subjectively determined, and do not lend themselves to tests of statistical significance. Such a rigorous point of view may well be in order when factoring is used as a means of testing different theories, or when weighted factor scores derived from one set of data are applied with spurious precision to another set. Even a strong proponent of factor analysis (Wrigley, 1958) has suggested that factor analysis as a method is worthwhile only if factors reported from study to study prove reasonably invariant. He adds that relative

invariance becomes possible only with the use of completely nonjudgmental, objective techniques.

The present results provide excellent examples of the degree of invariance that may be expected when using each of several objective factor analytic rotational procedures. They warn against basing an interpretation of a factor on too few items, even though the factor loadings may appear substantial. Note, for example, that several items with loadings approaching .50 and even .60 for one group of subjects have negligible loadings in the cor-

TABLE 4  
NORMALIZED VARIMAX ROTATION FOR GROUP 2

Variable	Factors <sup>a</sup>								h <sup>2</sup>
	I	II	III	IV	V	VI	VII	VIII	
1. Shaky-Steady	-19	14	06	-22	-20	67	25	-13	68
2. Strong-Weak	18	-13	-30	52	29	-17	01	06	53
3. Rugged-Delicate	-01	-09	-07	54	13	-08	-13	06	35
4. Sick-Healthy	-27	14	24	-02	-20	29	05	-13	29
5. Hard-Soft	-02	03	02	72	10	-05	-01	08	54
6. Heavy-Light	17	14	-82	07	01	01	10	02	74
7. Bad-Good	-14	65	-02	-02	04	14	39	11	63
8. Tired-Refreshed	-25	03	-01	-03	-15	63	-03	-10	49
9. Big-Little	-04	-12	-31	21	-22	-01	06	-11	22
10. Nice-Awful	05	-64	03	09	14	-19	-15	24	56
11. Masculine-Feminine	13	-22	-18	-07	07	17	-43	40	48
12. Underweight-Overweight	-01	-02	83	-01	-04	15	10	-07	73
13. Empty-Full	02	28	26	-33	-11	31	45	-07	57
14. Kind-Cruel	19	-66	10	-02	03	-16	-21	22	60
15. Strange-Natural	-23	16	25	-08	14	-07	58	-29	59
16. Useful-Useless	38	-38	14	12	09	-10	-60	-03	70
17. Confident-Unsure	49	-15	-03	33	26	-22	-11	22	55
18. Success-Failure	76	-31	-15	-05	12	01	-14	17	76
19. Dangerous-Harmless	-05	19	02	00	02	00	67	11	50
20. Happy-Sad	70	21	-10	05	14	-28	-07	02	65
21. Worthless-Valuable	-40	44	14	-14	-25	08	38	19	64
22. Stable-Unstable	62	-09	06	32	05	-24	-21	08	61
23. Destructive-Constructive	-09	21	-07	-09	-06	07	77	08	67
24. Active-Passive	31	-23	26	-08	29	-16	-41	14	52
25. Afraid-Unafraid	-64	11	03	-16	09	26	22	13	59
26. Oversexed-Undersexed	10	-04	-06	-07	-02	10	10	52	38
27. Friendly-Unfriendly	11	-50	09	00	13	-16	34	27	75
28. Embarrassed-At ease	-32	09	16	-19	-09	17	25	14	29
29. Loved-Hated	33	-72	01	03	03	11	-29	03	73
30. Satisfied-Dissatisfied	73	-12	-06	-21	12	-11	-08	09	64
31. Angry-Calm	-25	46	-03	19	04	31	38	08	56
32. Cold-Hot	22	37	-02	15	-20	25	06	05	32
33. Leader-Follower	-01	03	-27	09	57	-13	09	28	51
34. Sharp-Dull	18	00	16	32	60	08	-10	-02	54
35. Nervous-Relaxed	-17	20	20	-12	-08	79	-02	15	64
36. Superior-Inferior	11	-26	-04	22	53	-14	06	04	44
37. Smart-Dumb	26	-19	-01	06	70	-18	-16	-26	72
38. Popular-Unpopular	50	-43	-09	04	24	-08	05	-22	56
39. True-False	34	-49	08	11	27	-03	-40	-15	63
40. Shy-Outgoing	-01	16	07	03	-19	11	33	-02	19
Percentage of variance	10.7%	9.8%	5.4%	5.0%	5.8%	5.8%	9.2%	3.1%	54%

<sup>a</sup> Decimals omitted.



TABLE 5

INDICES OF FACTOR SIMILARITY: COMPARISON OF NORMAL VARIMAX ROTATIONS IN TWO EQUATED GROUPS<sup>a</sup>

Group 1 rotated factors	Group 2 rotated factors							
	I Success, Satisfied, Happy	II Loved, Kind, Good	III Overweight, Heavy	IV Hard, Rugged, Strong	V Smart, Sharp, Leader	VI Nervous, Shaky, Tired	VII Destructive, Dangerous, Strange	VIII Oversexed, Masculine
Success, I Stable, Valuable	84	59	13	38	47	-41	-53	25
Kind, II Nice, Good	56	81	-60	18	35	-31	-58	15
Heavy, III Overweight, Big	08	02	88	20	10	-24	-08	12
Strong, IV Hard, Rugged	50	-20	24	74	58	-43	-29	14
Leader, V Popular, Sharp	62	56	51	33	79	-37	-35	08
Afraid, VI Sick, Tired	-72	-49	-33	-33	-49	76	46	-09
Destructive, VII Embarrassed, Strange	-43	-51	-17	-28	-30	47	67	-09
VIII Oversexed	10	20	-04	-13	11	-10	-29	63

<sup>a</sup> Decimals omitted.

responding factor extracted from data for the equated subject group (see Items 25, 29, 37 in Tables 3 and 4). Some of these irregularities may be due to unreliability of the measure employed in the present study (Norman, 1959, has called attention to the unreliability of the semantic differential), and factoring "better" data might give more stable results. Nevertheless, they suggest the use of considerable caution in the interpretation of factor analytic results even where replication has been employed.

Despite the irregularities that have been noted, a general interpretation of the present results is possible. Encouragingly, there are broad similarities in all rotational solutions, despite such gross variations in technique as (a) utilizing both oblique and orthogonal rotations, (b) normalizing the data in one situation and not in another, and (c) rotating

seven factors in one instance and eight in the others.

*Identification of factors.* Factor I, Self-Confidence, is defined in all analyses by the items confident, successful, happy, valuable, stable, and satisfied. The negative pole suggests a depressed, deflated picture associated with frustration and failure. Factor I is very similar to an evaluative dimension called Self-Esteem which was isolated in an earlier study (Smith, 1958). It corresponds rather closely to the concept "acceptance of self" as described by Bills, Vance, and McLean (1951), and to Osgood's "evaluation" and "stability" dimensions as applied to the self (Osgood et al., 1957).

Factor II, Social Worth, is also highly evaluative in nature, and is similar to Osgood's factor of "social" and "moral" value. The moralistic implications of inherent good-

TABLE 6

INDICES OF FACTOR SIMILARITY: COMPARISONS OF BIQUARTIMIN ROTATIONS IN TWO EQUATED GROUPS\*

Group 2 rotated factors	Group 1 rotated factors						
	I Natural, Satisfied, Valuable	II Smart, Nice, Kind	III Heavy, Overweight, Big	IV Strong, Hard, Rugged	V Leader, Popular, Superior	VI Afraid, Sick, Unsure	VII Oversexed, Masculine, Angry
I Success, Satisfied, Happy	64	29	16	10	05	18	34
II Nice, Kind, Friendly	-12	66	-23	23	09	05	-06
III Natural, Overweight, Masculine	-10	23	50	-14	26	26	-08
IV Hard, Rugged, Strong	22	-13	-08	47	07	-27	14
V Superior, Big, Sharp	34	15	-03	16	53	-23	18
VI Nervous, Tired, Shaky	04	-04	-06	02	-10	43	17
VII Destructive, Dangerous, Useless	-06	-03	-09	08	-09	23	49

\* Decimals omitted.

ness or badness, and affective tones of hostility or love are conveyed. The items bad, awful, cruel, and unfriendly, and their opposites good, nice, kind, and friendly had high loadings in all analyses. More than Factor I, this dimension seems to tap one's judgment of inner worth or moral fiber.

Factor III, Corpulence, is almost exclusively defined by adjectives referring to physical size: big-little, overweight-underweight, heavy-light. It is clearly differentiated from Factor IV, Potency, which probably represents psychological as well as physical strength or power. In a previous analysis of the semantic differential (Smith, 1960) these two factors were not differentiated. Items from Factor IV (strong, rugged, hard) and Factor III (Corpulence) were grouped together to form a "body image" dimension. In the present subjects, self-ratings of strength and weakness tended to be somewhat independent of body size. This finding fails to support the

assumption that perceived body size is a function of strength and competence in mastering life's affairs.

Factor V, like the first two factors, has strong evaluative connotations. It appears to reflect subjectively perceived intelligence and leadership ability (superior, leader, sharp, smart) versus dulled, uncritical conformity (follower, dull, shy, inferior, dumb). It is named Independence after a similar factor isolated earlier (Smith, 1959, 1960).

Factor VI, Tension-Discomfort, is defined by items descriptive of anxiety and neurotic reactions often noted in psychiatric patients (nervous, tired, sick, shaky). Although correlated with the value factors (Factors I, II, IV, and V), this dimension may tap a more transient or immediate attitude toward the self, based on feeling states predominating at the time of rating. The writer has elsewhere speculated that such attitudes might be more easily subject to altera-

tion by tranquilizing and energizing drugs, counseling, or other therapeutic agents, and that changes on a dimension such as this may be involved when patients report symptomatic relief of acute personal distress.

There were two additional factors which were not clearly replicated. In the orthogonal solutions, one referred to sexual drive or masculine-feminine strivings; the other to feelings of estrangement, embarrassment, and loss of control. In the biquartimin solution these two dimensions were collapsed. It is possible that the inclusion of psychotic patients in the samples accounted for finding an estrangement or alienation factor.

All of the dimensions except Factor III have rather clear evaluative connotations. Osgood et al. (1957) have noted that the broad evaluative dimension they originally isolated can be factored into several specific value factors. While tests of self-concept in which items are treated additively to obtain an overall positive or negative score may be fairly parsimonious instruments for getting a quick, rough estimate of self-regard, they overlook the subtler multifactorial value structure of the self-concept. If ways can be found to eliminate the social desirability bias which usually enters into tests of this type, even a clearer picture of multidimensionality should emerge.

#### SUMMARY

Two equated groups of 96 subjects rated the self-concept on 40 semantic scales. Each set of data was separately factor analyzed and rotated by three objective procedures. Corresponding pairs of rotated solutions were compared to determine which method yielded greatest invariance. Kaiser's normal varimax method of rotation provided the most satisfactory factor structure for interpretation. Six dimensions of the self-concept were identified, and were called Self-Confidence, Social Worth, Corpulence, Potency, Independence, and Tension-Discomfort.

#### REFERENCES

- BARLOW, J. A., & BURT, C. The identification of factors from different experiments. *Brit. J. statist. Psychol.*, 1954, 7, 52-56.
- BILLS, R. E., VANCE, E. L., & MCLEAN, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, 15, 257-261.
- BOSTIAN, D. W., SMITH, P. A., LASKY, J. J., HOVER,

- G. L., & GING, R. J. Empirical observations on mental-status examination. *AMA Arch. gen. Psychiat.*, 1959, 1, 253-262.
- CARROLL, J. B. Biquartimin criterion for rotation to oblique simple structure in factor analysis. *Science*, 1957, 126, 1114-1115.
- CARROLL, J. B. Oblimin rotation solution in factor analysis: Computing program for the IBM 704. Harvard University, 1958. (Mimeo)
- GOLDBERG, L. R., & SMITH, P. A. The clinical usefulness of the Archimedes spiral in the diagnosis of organic brain damage. *J. consult. Psychol.*, 1958, 22, 153-157.
- HARMAN, H. H. *Modern factor analysis*. Chicago, Ill.: Univer. Chicago Press, 1960.
- KAISER, H. F. An analytic rotational criterion for factor analysis. *Amer. Psychologist*, 1955, 10, 438. (Abstract)
- KAISER, H. F. Note on Carroll's analytic simple structure. *Psychometrika*, 1956, 21, 89-92.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.
- KUMATA, H., & SCHRAMM, W. A pilot study of cross-cultural meaning. *Publ. opin. Quart.*, 1956, 20, 229-238.
- LASKY, J. J., HOVER, G. L., SMITH, P. A., BOSTIAN, D. W., DUFFENDACK, S. C., & NORD, C. L. Post-hospital adjustment as predicted by psychiatric patients and by their staff. *J. consult. Psychol.*, 1959, 23, 213-218.
- NEUHAUS, J. O., & WRIGLEY, C. F. The quartimax method, an analytic approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, 7, 81-91.
- NORMAN, W. T. Stability-characteristics of the semantic differential. *Amer. J. Psychol.*, 1959, 72, 581-584.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.
- SMITH, P. A. Factor analytic studies of the self-concept. Paper read at Midwestern Psychological Association, May 1958.
- SMITH, P. A. A factor analytic study of the self-concept in psychiatric patients. Extended Report SP-108, September 1959, System Development Corporation.
- SMITH, P. A. A factor analytic study of the self-concept. *J. consult. Psychol.*, 1960, 24, 191.
- SMITH, P. A., BARROWS, H. S., & WHITNEY, J. N. Psychological attributes of occupational therapy crafts. *Amer. J. occup. Ther.*, 1959, 13, No. 1.
- TUCKER, L. R. A method for synthesis of factor analysis studies. Personnel Research Section Report No. 984, 1951, Department of the Army.
- WRIGLEY, C. F. Review of Y. Ahmavaara, On the unified factor theory of mind. *Contemp. Psychol.*, 1958, 3, 184-185.
- WRIGLEY, C. F., & NEUHAUS, J. O. The matching of two sets of factors. Contract Memorandum Report A-32, 1955, University of Illinois.

(Received January 17, 1961)



## IDENTIFICATION, AUTHORITARIAN DEFENSIVENESS, AND SELF-ESTEEM<sup>1</sup>

EZRA STOTLAND AND MAX L. HILLMER, JR.

*University of Washington*

This paper reports one of a series of studies of the process of identification, which employs the following cognitive formulation of identification: When an individual conceives of himself and some other person as having some trait, characteristic, or attribute in common and further conceives of this other person as having some additional trait associated with their common trait, then the first individual will tend to conceive of himself as also possessing the second trait. This tendency is qualified by pre-existing conceptions by the individual of his possessing attributes contradictory to, or incompatible with, the second attribute. That is, the individual sees as co-existent or contiguous two attributes in another person, one of which he perceives himself as having (called the FSA or first similar attribute). He will then tend to conceive of himself as sharing the second attribute (called the DSA or derived similar attribute) unless there is some barrier to the conception. The assumptions underlying the formulations are similar to Tolman's (1949) notion of field expectancies.

The first implication of this formulation is that no interaction or communication relationship need exist between the individual and the person who serves as the identification model. Secondly, the trait which the individual comes to accept as his own need not be desirable nor rewarding in itself. A third implication is that the relation of the FSA and the DSA may be entirely arbitrary. In earlier studies in this series (Burnstein, Stotland, & Zander, 1961; Stotland & Patchen, 1961; Stotland, Zander, & Natsoulas, 1961) it was possible to interpret the relationship between FSA and DSA as meaningful or logical. That is, the subject might have acquired the DSA only because it seemed reasonable that a person having the FSA should have

the DSA also. The present study is designed to demonstrate that the process does occur when the relationship of FSA and DSA is entirely arbitrary and even when the possession of the FSA by the subject is simply a matter of chance.

A fourth point that is explored in this study is whether the individual's degree of involvement in the FSA will influence his tendency to identify. In the study by Stotland et al. (1961) it was found that the persons with a high degree of interest in the FSA identified more. The present study provides an opportunity to test the relationship again.

A fifth point that is explored in this study is whether other relationships besides similarity can have an influence on an individual's conception of himself. Instead of perceiving an FSA between himself and a model, a person may perceive that the model has some attribute opposite to one he perceives himself as having. By opposite is meant involving processes which are contrary to, or negations of, each other (e.g., good vs. bad; up vs. down). Oppositeness is to be distinguished from low degree of similarity or difference. In oppositeness, the two attributes are relevant to each other, that is, in some common dimension. In difference, the attributes are not relevant. The distinction between oppositeness and difference is analogous to Kelly's (1955) contrast between the poles of a dichotomy and the range of convenience of a construct. The question explored in the present study was whether oppositeness generalized, as well as similarity. When the person and the model are opposite in some attribute (A), will the subject conceive himself as opposite to the model with respect to some other attribute associated in the model with the first attribute (A)? Such generalization of oppositeness would be akin to the concept of negative identification (Newcomb, 1950).

<sup>1</sup> This research was conducted under a grant from the National Institute of Mental Health (M 2433).

A sixth consideration is the question of individual differences in the tendency to identify. Two aspects of personality suggested themselves as relevant to the lability of the self-concept and to susceptibility to social influence. The first of these was self-esteem. Hovland and Janis (1959) and Stotland, Thorley, Thomas, Cohen, and Zander (1957) have found persons with low self-esteem are more easily influenced. We might hypothesize that the low self-esteem person has a self-concept that offers less resistance to the process of identification than the high self-esteem person.

Stotland and Patchen (1961) found that persons low in authoritarian defensiveness identify more readily than high authoritarians. A possible interpretation of this finding is that high authoritarian has a rigid self-concept while the low authoritarians have less need to defend themselves by keeping others at a distance. This is consistent with Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) where the low authoritarian is described as warmer and more empathetic. This study further explores the relationship between these two personality characteristics that may affect the individual's tendency to conceive of himself as related to others.

A final problem explored in the present study is whether the DSA remains at the level of an individual's perception of himself, or whether it will also be manifest in the overt behavior indicated by the DSA.

The hypotheses of the present study are as follows:

**Hypothesis 1:** The first hypothesis concerns an individual who is made to perceive that he is similar to a model with regard to a given attribute (FSA). The more of another attribute the model is perceived to possess, the more will the individual perceive himself to possess this attribute.

**Hypothesis 2:** If an individual is made to perceive that he is opposite to a model with regard to a given attribute, he will perceive that he has the opposite of some other attribute of the model which is associated with the first attribute.

**Hypothesis 3:** The processes described in Hypotheses 1 and 2 will occur more if the

person is highly involved in the FSA than if he has low involvement.

**Hypothesis 4:** Persons of low self-esteem will show the effects predicted in Hypotheses 1 and 2 more than persons high in self-esteem.

**Hypothesis 5:** Persons with low scores on the Defensiveness subscale of the F Scale will show the effects predicted in Hypotheses 1 and 2 more than persons high on the subscale.

## METHOD

**Overview.** The data were collected during two regular sessions of the senior author's class in introductory psychology. The students had no advance notice of the experiment other than the understanding that they would participate in some classroom experiments during the quarter. At the first session the subjects filled out the Defensiveness subscale of the F Scale and a measure of self-esteem. Several weeks later the class filled out the experimental booklet. No relationship between these two tasks was pointed out to the students.

**Subjects.** The subjects were 92 students, 37 female and 55 male, in an introductory psychology class at the University of Washington. They were predominantly freshmen and sophomores.

**First session.** A booklet containing a subscale of the F Scale and a modification of the Q sort to measure self-esteem was administered. McClintock (1958) found it useful to divide the F Scale into an Other-Directedness and a Defensiveness subscale. The scale used here is essentially McClintock's Defensiveness scale. It is comprised of items which were categorized under the headings of authoritarian aggression, anti-intraception, superstition, and stereotype, destructiveness and cynicism, projectivity, and sex (Adorno et al., 1950).

The modification of the Q sort, based on the work of Weinberger (1951) consisted of a check list of 26 desirable traits—adjectives, presented first with instructions to check how important it was to stand high in each trait, and next (on a separate page) with instructions to indicate how high they actually did stand on each trait. Each trait was checked on a five-point scale with the restriction that only five traits were to be placed in any one category except for six in the middle category. The self-esteem score was the sum of the squared differences in rating for each adjective between the two administrations, the greater the sum, the less the self-esteem.

**Second session.** The booklet administered at the second session was designed to give each subject an experience which would serve as a basis for identification (FSA), to present verbally an identification model, and finally to determine if identification occurred.

The first sheet of the booklet described the experiment as part of a continuing program by the psychology department to assess the suitability and interest value of a great variety of psychological tests



for use with college students. For a full page these instructions hammered home the point that each booklet contained a random sample of these tests, that no two subjects would be likely to have the same tests and that the particular assortment of tests any student received was simply a matter of chance. (The experimenter, in his verbal instructions, re-emphasized these points.) The purpose of these instructions was to provide a plausible rationale for the experiment and to create the illusion that each subject was having a unique experience. The purpose of having the subjects work on the tasks was to give each one an experience which could later function as the basis of an FSA with a model. By using tasks for the purpose, the subject was forced to attend to the tasks with at least minimal involvement in them.

The next four pages of the booklet consisted of the four "tests." Although each booklet contained the same four tests, the appearance and order of the tasks was varied so that even if a subject violated instructions and glanced at his neighbor's paper he would see a "different" task. This difference was obtained by using different spatial arrangements on the pages, using four different type faces for four versions of each task, using various colors of paper, dittoing two versions of each task and mimeographing two versions and finally assembling these tasks in an order that would prevent any two identical booklets from occurring side by side. The tasks used had been found in pilot studies to be fairly high in interest value to college students, yet were tasks which almost all could complete without error or knowledge of error. These tasks were also selected because it would be possible later to describe tasks opposite to each. The tasks were:

1. *Finding causes*, in which a noun was presented and the subject chose the cause of that noun from among a list of four other nouns. Twelve such cause-effect items were employed.
2. *Finding antonyms*, in which the subject had two lists of 12 words. He would then match each and form one list with its opposite in the other list.
3. *Letter-number substitution*, in which the subject translated a simple sentence into numbers according to a code presented to him.
4. *Word formation*, in which the subject was presented with a list of nine sets of three short words and another list of nine long words. Each of nine long words had been formed from one set of the three shorter ones. On this basis, the subjects had to match these two lists.

To measure the subject's degree of involvement in the task and to make the rationale for giving the tasks more plausible, each subject rated his interest in each task on a nine-point rating scale.

After these tasks came a sheet which presented the identification model. The subjects read that this part of the study was to ascertain the impressions people form of another person when they have only a limited amount of information about that person. This paragraph was the same for all conditions.

The second paragraph went on to say that a person assigned the code "Pat" had participated in this same study the preceding year. To establish the Similarity condition the paragraph went on to say that Pat had taken exactly the same four tests as the subject had just completed, and proceeded to enumerate them. This served to establish the FSA, an identical experience. For the Opposite condition the paragraph described Pat as having worked on exactly opposite tasks, where the subject had worked on antonyms, Pat had worked on synonyms; instead of finding causes, Pat found effects; instead of finding one long word consisting of three short ones, Pat found three short words from one long word. For these subjects, then, the FSA was performance of tasks that were relevant but at the opposing end of a common dimension. For the Control condition this paragraph was omitted entirely.

The next paragraph served to describe a second attribute of Pat, which constituted the attribute which the subject could make into a DSA by accepting it into his self-concept, that is, by identifying with the model. The subjects read that, in another study, entirely unrelated to the present one, Pat took a "complex, detailed task which requires both speed and accuracy in handling several factors at the same time." Half of the sheets described Pat as performing considerably above average and half as considerably below average. In this way half of the subjects in each of the three conditions pertaining to the FSA received a sheet describing Pat favorably on the DSA and half received an unfavorable description.

A final paragraph, the same for all subjects, asked them to indicate on a five-point scale how interested they would be in meeting Pat if they were permitted to talk about anything except the experiment. This provision about not talking about the experiment was introduced to prevent the identity of the experiences of the model and the subject in the Similarity condition from giving the latter more incentive than subjects in other conditions to talk with the model.

The next page in the booklet was devoted to an instruction sheet for a specially constructed task designed to reveal the subjects' acceptance of Pat's level of performance into their own self-concept. They were told to read the instructions and examples and to await verbal instructions from the experimenter before beginning the test as it was to be timed.

Although no explicit reference was made to the "complex, detailed task" which Pat was alleged to have taken, anyone reading the instructions could see that such a description would be quite appropriate to the current task. The task consisted of two pages of 350 words each, arranged in five columns to the page. The subject was to underline every word which met certain criteria. One criterion, for example, was to underline any word which contained both the letters "a" and "u" in which the "a" appeared earlier in the word than the "u." This task was adopted after an examination of a wide range of ability tests because individual differences in performance on it



seemed relatively less dependent on intelligence or clerical aptitude than on motivational or attitudinal factors.

Before beginning the task the subjects were asked to indicate on a nine-point rating scale how well they thought they would do on the task, in this way providing an index of their acceptance of Pat's ability level as their own. At the completion of the task they were asked to indicate on a nine-point scale how well they thought they had performed and on a third scale to indicate how they thought they would do on this or a similar task if they were to take it in the future. These provided three measures of the subject's acceptance of Pat's ability level.

Originally it had been hoped that actual differences in performance level could be used as a DSA, but the instructions for the task were inadequate so that speed and accuracy scores were unusable.

The complete experimental booklet thus consisted of a cover (instruction) sheet followed by a section of four short tests, identical in content but varied in order and appearance to make four forms, A, B, C, and D. The next sheet of the booklet, describing Pat, had six forms corresponding to the six conditions of the experiment.<sup>2</sup> The six conditions were Similar-Good, Similar-Poor, Opposite-Good, Opposite-Poor, Control-Good, Control-Poor. If we designate these six sheets as a through f we can see that there are now 24 forms of the booklet, Aa, Ab-Af through De, Df. All the booklets ended with the four sheets comprising the underlining task with its instructions and rating scales.

The booklets were then carefully "stacked" in an Aa, Be, Cc, Df, Ab, Bd, Ca, Da, etc. order. This served to insure that each of the six conditions was equally represented when the booklets were distributed in class and also insured that the appearance of adjacent booklets was unlike.

The subjects were told that the booklet that each received contained a different random sample of tasks from a large number of tasks. The booklets were then distributed to the subjects just as the former were stacked, the particular booklet the subject received being a chance result of where they happened to sit.

At a later date the purpose of the experiment was discussed with the class and questions concerning it were answered.

## RESULTS

Table 1 presents the mean self-evaluation indices, obtained by summing each subject's three ratings, for each of the six experimental

<sup>2</sup> In actuality, there were two other conditions employed, in which the subjects read that Pat had taken different tests and that he was either good or poor in the DSA task. However, a typographical error invalidated these conditions. Since the subjects who received these two types of booklets were randomly selected, the error only has the effect of reducing the total Ns in the study.

TABLE 1  
MEAN SELF-EVALUATION INDEX FOR SUBJECTS IN EACH EXPERIMENTAL CONDITION (Ns in parentheses)

Pat's level of ability	Cognitive relationship to the model		
	Similar	Opposite	Control
Good	a 13.00 (17)	c 15.67 (15)	e 13.31 (16)
Poor	b 16.21 (14)	d 13.37 (15)	f 13.40 (15)

Note.—Cells significantly different: a vs. b ( $t = 3.41$ ,  $p < .001$ , one-tailed), a vs. c ( $t = 2.02$ ,  $p < .03$ , one-tailed); b vs. d ( $t = 1.81$ ,  $p < .05$ , one-tailed). The lower the score, the better the evaluation.

conditions. Table 2 shows the results of an analysis of variance for these data. Only the interaction  $F$  ratio was significant. The locus of this interaction is investigated in Table 1 by means of  $t$  tests.

As predicted by the first hypothesis, subjects in the Similar-Good condition rated themselves significantly better than those in the Similar-Poor condition. In the Opposite condition the difference that occurs is in the direction that the second hypothesis predicts but it fails to reach significance. The lack of a difference in the Control condition is as expected, since these subjects lacked the FSA.

The significant differences between the Similar- and Opposite-Good conditions and between the Similar- and Opposite-Poor conditions are consistent with the joint functioning of the first and second hypotheses. Thus while the second hypothesis is not directly confirmed, some data in Table 1 suggest support for it.

TABLE 2  
ANALYSIS OF VARIANCE OF SELF-EVALUATION INDEX

Source	SS	df	MS	F
1. Cognitive relationship to model	1.00	1	1	
2. Level of ability of Pat	55.00	2	27.5	
3. Interaction	104.00	2	52	3.939*
4. Within cells <sup>a</sup>	1028.00	78	13.2	

<sup>a</sup> The Ns in the cells were equalized to 14 for the analysis by removing the appropriate number of subjects from each cell who were closest to the mean of that cell, thus magnifying the estimate of error variance and retaining the more reliable estimate of the cell means, based on the larger N.

\*  $p < .05$ .

TABLE 3

MEAN SELF-EVALUATION INDEX FOR SUBJECTS IN EACH  
EXPERIMENTAL CONDITION HIGH AND LOW  
INTEREST ON THE FIRST  
FOUR TASKS  
(*N*s in parentheses)

Degree of interest in first four tasks	Pat's level of ability	Cognitive relationship to the model		
		Similar	Opposite	Control
High	Good	a 12.28 (7)	c 17.29 (7)	e 13.80 (10)
	Poor	b 16.25 (4)	d 13.10 (10)	f 9.00 (6)
Low	Good	g 13.50 (10)	i 14.25 (8)	k 12.50 (6)
	Poor	h 16.20 (10)	j 13.00 (5)	l 18.38 (8)

Note.—Difference between mean differences: (e - f) vs. (k - l) ( $t = 3.20$ ,  $p < .01$ , two-tailed, Walker & Lev, 1953). Significant differences between cells:

## HIGH INTEREST SUBJECTS

- a vs. b ( $t = 1.65$ ,  $p < .07$ , one-tailed)  
c vs. d ( $t = 2.19$ ,  $p < .05$ , one-tailed)  
a vs. c ( $t = 2.27$ ,  $p < .03$ , one-tailed)  
e vs. f ( $t = 3.24$ ,  $p < .01$ , two-tailed)  
b vs. f ( $t = 3.03$ ,  $p < .01$ , one-tailed)  
c vs. e ( $t = 1.72$ ,  $p < .06$ , one-tailed)  
d vs. f ( $t = 2.95$ ,  $p < .01$ , one-tailed)

## LOW INTEREST SUBJECTS

- j vs. h ( $t = 1.97$ ,  $p < .06$ , two-tailed)  
k vs. l ( $t = 1.92$ ,  $p < .07$ , two-tailed)

Overall inspection of the cell means shows that only the Similar-Poor and Opposite-Good conditions have means which depart from the neighborhood of 13. In these two cells, which produced the significant interaction, subjects rated themselves *worse* than in the other conditions.

The third hypothesis concerns the degree of involvement of the subjects with the FSA. In an index of involvement the subjects' ratings of the interest value of each of the first four tasks was used. Each subject's ratings were summed and the total distribution split at the median to define the high and low interest groups. Table 3 shows the mean self-evaluation indices for the six conditions after this split. Analysis of variance was not employed in dealing with these data because equalizing all *N*s would have meant the loss of about half of the subjects. Recourse was

had to the method of Walker and Lev (1953, p. 158ff.) utilizing a  $t$  test of the difference between the mean differences of *high* and *low* interest groups in the Good and Poor conditions.

This test was first used to see if there was a significant difference between the high and low interest groups in the respective differences between Good and Poor conditions. No significant difference was found but significant differences between the Good and Poor conditions were obtained for both levels of interest.

When the Walker-Lev (1953) technique was applied to the oppositeness conditions, no significance was found between interest levels, but only in the high interest group was there a significant difference between the Good and Poor condition. This suggests that a high level of interest favors the generalization of oppositeness. A significant difference between Cells a and c suggests the joint functioning of the generalization of similarity and oppositeness.

Results involving the control condition are completely unexpected. On the basis of this theory no differences between Good and Poor

TABLE 4

MEAN SELF-EVALUATION INDEX FOR SUBJECTS IN EACH  
EXPERIMENTAL CONDITION HIGH AND LOW  
IN SELF-ESTEEM  
(*N*s in parentheses)

Level of self-esteem	Pat's level of ability	Cognitive relationship to the model		
		Similar	Opposite	Control
Low	Good	a 13.63 (8)	c 15.75 (8)	e 13.78 (9)
	Poor	b 20.65 (4)	d 13.33 (6)	f 16.14 (7)
High	Good	g 12.44 (9)	i 15.50 (6)	k 12.71 (7)
	Poor	h 14.60 (10)	j 14.00 (9)	l 10.67 (6)

Note.—Difference between mean differences: (a - b) vs. (g - h) ( $t = 2.10$ ,  $p < .025$ , one-tailed, Walker & Lev, 1953). Significant differences between cells: Low self-esteem subjects—  
a vs. b ( $t = 2.26$ ,  $p < .025$ , one-tailed); High self-esteem—  
none.

conditions were anticipated on either interest level. All comparisons in the control column were significant. Although the high interest control-good subjects were about average in their self-evaluations (higher than the Opposite-Good subjects, for example) the Control-Poor subjects rated themselves unusually high, so Cells b and d also differ significantly from Cell f. In the low interest group a reversal occurs so that the Control-Good subjects rate themselves favorably and the Control-Poor subjects' self-evaluations are markedly depressed. Perhaps the low interest group had a somewhat passive and indifferent attitude toward the tasks and merely adopted Pat's ability level as a convenient yardstick by which to estimate their own performance whereas the high interest group had a rather competitive attitude and took Pat's low performance as a challenge. Whatever the explanation, it seems likely in retrospect that the interest scales are an index to some undefined attitudes toward the tasks, and possibly the whole experiment, rather than an indicator of what was conceived of originally as "involvement." The time, effort, and activity of each subject in performing the pretests

TABLE 5  
MEAN SELF-EVALUATION INDEX FOR SUBJECTS IN EACH  
EXPERIMENTAL CONDITION HIGH AND LOW  
IN AUTHORITARIAN DEFENSIVENESS  
(Ns in parentheses)

Level of authoritarian defensiveness	Pat's level of ability	Cognitive relationship to the model		
		Similar	Opposite	Control
Low	Good	a 11.70 (10)	c 15.71 (7)	e 13.17 (6)
	Poor	b 16.50 (8)	d 12.56 (9)	f 14.11 (9)
High	Good	g 14.86 (7)	i 15.63 (8)	k 13.40 (10)
	Poor	h 15.83 (6)	j 15.50 (6)	l 12.33 (6)

Note.—Difference between mean differences: (a - b) vs. (g - h) ( $t = 1.65$ ,  $p < .06$ , one-tailed, Walker & Lev, 1953). Significant differences between cells: Low defensive subjects— $a$  vs.  $b$  ( $t = 3.24$ ,  $p < .01$ , one-tailed),  $c$  vs.  $d$  ( $t = 1.99$ ,  $p < .05$ , one-tailed); High defensive subjects—none.

TABLE 6

MEAN RATINGS\* OF DESIRE TO MEET THE MODEL IN  
EACH OF THE SIX EXPERIMENTAL CONDITIONS  
(Ns in parentheses)

Pat's level of ability	Cognitive relationship to the model			
	Similar	Opposite	Control	Totals
Good	a 3.00 (17)	c 2.88 (16)	e 2.50 (16)	g 2.70
Poor	b 3.57 (14)	d 3.40 (15)	f 3.35 (14)	h 3.44
Total	3.26	3.13	2.90	

Note.—Cells significantly different by two-tailed  $t$  test:  $e$  vs.  $f$  ( $t = 2.690$ ,  $p < .01$ );  $g$  vs.  $h$  ( $t = 3.719$ ,  $p < .01$ ).  
\* The lower the rating, the greater the desire.

probably "involved" the subjects above whatever the threshold of involvement may be for identification to occur.

The fourth hypothesis was that subjects low in self-esteem would conform to the similarity and opposite hypotheses more than high self-esteem subjects. The total distribution of self-esteem scores was dichotomized at the median and means for the six conditions are presented in Table 4. Both the Walker-Lev (1953) test and the difference between low self-esteem Good and Poor conditions were significant in the similarity column, showing that low self-esteem persons identify more than highs. The self-evaluation index for the low self-esteem, Similar-Poor subjects is over 4 points lower than any other mean in the table. It appears that the low self-esteem subjects are particularly vulnerable to identifying with negative attributes.

The fifth hypothesis is that identification will occur more readily in subjects low in authoritarian defensiveness than those high in this characteristic (Table 5). The Walker-Lev (1953) test indicated a difference between the high and low authoritarians in the similarity condition. In both the Similar and Opposite conditions for the low defensiveness group the Good and Poor conditions differed significantly. Thus low defensives identify more and give some evidence of generalizing oppositeness more.

Since both the authoritarian defensiveness scale and the self-esteem scale had some



value as predictors their correlation was calculated and found to be .02.

Another point investigated was the effect of attraction to the model on readiness to identify. Studies by Back (1951) and Raven (1959) suggest that attraction would lead to pressure to conform to the model. The hypothesis of the present study is that identification is a cognitive process, with perception of similarity between self and model as the necessary condition. Motivation to interact with the model is not assumed to be relevant.

The subjects' ratings of their desire to converse with the model provides evidence on this point. Table 6 presents the mean ratings of the subjects' attraction to Pat. There was an overall preference for the "good" model over the "poor," a result in perfect consonance with common sense. This preference is manifest within all conditions, although it is significant only in the Control condition. Thus while differences in attraction occurred in all conditions, only in the Similarity condition did the subjects acquire the DSA. Thus it appears that identification and attraction can vary independently.

### DISCUSSION

One of the major purposes of the present study was to determine whether an individual will identify with another person even when the FSA is a purely arbitrary one. In the earlier studies both meaningful and arbitrary relationships between the FSA and DSA may generate the results. In the present study, arbitrariness was operationalized in two ways: first, the subjects were explicitly told that the assignment of tasks on which the FSA is based was purely arbitrary, and the various booklets containing the tasks were very obviously assigned to them in a random fashion. Secondly, the association between the FSA and the DSA was purely arbitrary, since the model's having taken the same tasks as the subject was also a matter of chance. The results indicated the generalization of arbitrary similarity postulated to underlie identification.

The arbitrariness of the subjects' reactions does not, however, imply that they were behaving in a blind, completely irrational fashion. It implies only that the generalization of

similarity does not appear to be limited to cases where such generalization would be supported by syllogistic logic. What the subjects may be doing is attempting to maintain a consistent, stable perception of the world, even if this requires a little logical sloppiness. The consistency of the person's picture of the world is maintained by generalizing similarity, even when the generalization is logically unwarranted. As will be suggested below, the reactions to the experiment of persons low in self-esteem is also understandable in terms of attempts to maintain consistency.

The present approach also contends that identification will occur even when the DSA has no value to the person. This contention is supported by the finding that the results for the total sample were generated primarily by the subjects' poor self-evaluation in the Similar-Poor condition.

The present theory would also hold that identification can be understood in terms of self-perception and perception of other persons and that communication or interaction between the person and the model is not a requisite. The results are consistent with this contention, since the total sample of subjects desired more to talk with the good model than with the poor. Yet, only when the subjects had an FSA between themselves and the model did they tend to acquire the model's attributes.

The hypothesis that persons of low self-esteem would identify more was confirmed, although they did not tend to generalize oppositeness more. However, the pattern of results raises questions about the rationale given in the theoretical discussion for expecting low self-esteem persons to identify more. There, it postulates that low self-esteem persons are more susceptible to any social pressure. In the present study, however, they reacted mostly by very low self-evaluation in the Similarity-Poor condition, rather than showing as well the high self-evaluation in the Similarity-Good condition that would be expected if the reaction were based simply on vulnerability to social pressure. One possible interpretation of the present results is that persons of low self-esteem can accept new attributes into their self-concepts which

are consistent with the general evaluation they have of themselves. In the theoretical discussion, it was pointed out that potential DSAs which are inconsistent with already existing attributes of the self-concept would not be acquired through identification. Perhaps DSAs of high ability would be inconsistent with other attributes of the self-concepts of low self-esteem persons. This argument raises the interesting question of why the persons with high self-esteem did not have very high evaluations in the Similarity-Good condition. Perhaps, the originally present theory based on differential susceptibility to social influence has to be integrated with the argument based on consistency of the attributes of the self-concept.

The hypothesis was also confirmed that persons of low authoritarian defensiveness would identify more. There was also some indication that these persons also generalized oppositeness more. These findings are consistent with the picture of the low authoritarian personality as being more empathic and warm, better able to identify with others. One clue as to the reason for the difference in reaction between high and low authoritarian defensives comes from the data that the highs tend to perceive other persons as being like themselves (Crockett & Meidinger, 1956; Scodel & Mussen, 1953) and that they tend more to employ the defense mechanism of projection (Peak, Muncy, & Clay, 1960). These persons may have a strong tendency to generalize from themselves to other persons. This tendency may be sufficiently strong to override any tendency to identify, which entails generalization from others to oneself. Other relevant characteristics of the high authoritarians might be their rigidity and anti-intracception, which together would make them resist any changes in their self-conceptions, since they would be averse to change and would not be prone to subject their self-concepts to observation and change.

More generally, the present study indicates the possibility of taking so intangible a Freudian concept as identification into the experimental laboratory. This transfer entails a certain simplification, such as reducing it to Tolmanian terms of sign-significate. However, this simplification has the value of show-

ing the power of a theory like Tolman's to encompass a wide variety of psychological phenomena. Furthermore, the cognitive flavor of this simplification can lead to other fruitful hypotheses, such as the generalization of oppositeness which would not have been derived easily from the original Freudian formulation.

This theory was based on a cognitive theory of generalization of similarities. The very statement of the theory would raise the question of whether other cognitive processes would not occur in the interaction between an individual's conception of himself and his conception of other persons. One such cognitive process explored in the present study is that of the generalization of oppositeness, the complement to the generalization of similarity. The results indicate some support for the hypothesis that oppositeness generalizes. This hypothesis tended to be supported more by subjects highly involved or interested in the characteristics upon which the oppositeness is based. On the other hand, subjects of both high and low involvement in these tasks generalized similarity to the model. It appears that such generalization of oppositeness, which may be termed negative identification, is a more tenuous process than the generalization of similarity and, therefore, requires more involvement from the individual identifier. This relative weakness of the tendency to generalize oppositeness may be a reflection of the fact that generalization is a more primitive process developmentally. Another reason for the greater difficulty in demonstrating generalization of oppositeness than similarity might be that oppositeness is a more difficult concept to grasp and may not be distinguished by some persons from difference. The variable relevant to individual differences with regard to reactions to oppositeness might not have been tapped in the present study. In any case, when negative identification or the generalization of oppositeness does occur, it tends to have effects diametrically opposite to those of similarity identification.

#### SUMMARY

Subjects assembled in a large group were randomly assigned "different" sets of four tasks, and the random assignment was made

conspicuously visible to the subjects. After working on these tasks, they either read about another person who ostensibly had previously worked on the same set of tasks, or about someone who ostensibly had worked on an "opposite" set of tasks, or read nothing about this person having worked on any tasks. Half the subjects reading each of these three types of descriptions of this person were told that this person had been found to be good at clerical tasks. The other half read that he was poor at clerical tasks. The subjects were then given a clerical task, asked to predict how well they would do and to evaluate their performance on the task. It was found that:

1. Subjects who read the model was good in clerical tasks evaluated their performances more favorably than those who read that the model was poor, provided that the model had worked on the same tasks as themselves. This was especially true if the subjects were low in self-esteem or low in authoritarian defensiveness.

2. If the subjects read that the model had worked on tasks opposite to their own, there was some tendency for them to evaluate their performances less favorably when the model was good than when he was poor, especially if the subjects were very interested in the first four tasks given them, and especially if they were low in authoritarian defensiveness.

3. If the subjects received no information about the models' having worked on tasks similar to their own tasks, the subjects' evaluations of their performances on the clerical task were unaffected by the model's level of ability. However, the level of interest in the first four tasks appeared to have complex interactions with the effect of the model on the subjects' evaluations of performance.

4. The level of attraction to the model was found to vary independently of the effects of the experimental conditions.

The results were interpreted as supporting a cognitive theory of identification defined as the generalization of interpersonal similarity.

#### REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, 46, 9-23.
- BURNSTEIN, E., STOTLAND, E., & ZANDER, A. Similarity to a model and self-evaluation. *J. abnorm. soc. Psychol.*, 1961, 62, 257-264.
- CROCKETT, W. H., & MEIDINGER, T. Authoritarianism and interpersonal perception. *J. abnorm. soc. Psychol.*, 1956, 53, 378-380.
- HOVLAND, C. I., & JANIS, I. (Eds.) *Personality and persuasibility*. New Haven: Yale, 1959.
- KELLY, G. A. *Psychology of personal constructs*. Vol. I. New York: Norton, 1955.
- MCCLINTOCK, C. Personality syndromes and attitude change. *J. Pers.*, 1958, 26, 479-493.
- NEWCOME, T. M. *Social psychology*. New York: Dryden, 1950.
- PEAK, HELEN, MUNCY, BARBARA, & CLAY, MARGARET. Opposites structures, defenses, and attitudes. *Psychol. Monogr.*, 1960, 74(8, Whole No. 495).
- RAVEN, B. Social influence on opinions and the communication of related content. *J. abnorm. soc. Psychol.*, 1959, 58, 119-128.
- SCODEL, A., & MUSSEN, P. Social perception of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, 48, 181-184.
- STOTLAND, E., & PATCHEN, M. Identification and changes in prejudice and in authoritarianism. *J. abnorm. soc. Psychol.*, 1961, 62, 265-274.
- STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem on self-evaluation. *J. abnorm. soc. Psychol.*, 1957, 54, 55-63.
- STOTLAND, E., ZANDER, A., & NATSOULAS, T. Generalization of interpersonal similarity. *J. abnorm. soc. Psychol.*, 1961, 62, 250-256.
- TOLMAN, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, 56, 144-155.
- WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.
- WEINBERGER, B. Achievement motivation and self-concept. Unpublished honors thesis, University of Michigan, 1951.

(Received February 15, 1961)



## COMMUNICATION OF AFFECT IN "COOPERATIVE CONDITIONING" OF RHESUS MONKEYS<sup>1</sup>

ROBERT E. MILLER, JAMES H. BANKS, JR., AND NOBUYA OGAWA<sup>2</sup>

*School of Medicine, University of Pittsburgh*

An approach to the investigation of variables in the communication of affects was made feasible by the development of techniques for conditioning monkeys to react to each other in response to experimentally induced affective states (Murphy, Miller, & Mirsky, 1955). It was found that these affective experiences had profound effects on subsequent social relationships between the animals (Miller, Murphy, & Mirsky, 1955; Murphy & Miller, 1956). The behavior of these animals implied that very subtle emotional expressions were detected and interpreted in the constant and complex interplay in the experimental social situation. In an attempt to measure this communication of affect, a new method was developed (Mirsky, Miller, & Murphy, 1958) and some of the relevant expressive cues were identified (Miller, Murphy, & Mirsky, 1959a, 1959b).

The present report deals with a new and more direct approach to the study of the phenomena of communication of affects. While the former method measured the effect of such communication on an existing conditioned response, it did not require the animals to communicate and, therefore, was subject to extinction upon repeated testing. The present method, however, requires successful communication between the animals for mutual solution of an avoidance problem.

### METHOD

#### Subjects

Three postadolescent male rhesus monkeys were used in this experiment. Each of them had been sub-

<sup>1</sup> This investigation was supported by research grants from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service, the Foundations' Fund for Research in Psychiatry, and the Commonwealth of Pennsylvania.

<sup>2</sup> Nobuya Ogawa is a visiting Research Fellow from the First Department of Medicine, Kyushu University School of Medicine, Fukuoka, Japan.

jects in a previous investigation involving discriminated avoidance to visual form stimuli.

#### Apparatus

Standard equipment consisting of primate chairs, response bars, relay and timing panels, programmers, and cumulative recorders were used in these experiments. The only modification which was made was the insertion of a milk glass panel in front of the lights in a visual stimulus panel. This was done to reduce the intensity and diffuse the lights so that glare or reflections from the plastic and/or metal fittings of the primate chairs would not occur. The final result was effective in that the experimenter could not detect any reflection from in front of the chair using a violet stimulus light which was readily apparent from the monkey's position.

The animals were housed and tested in an air-conditioned isolation room. Two one-way vision screens were mounted in the walls enabling the experimenters to make observations of the animals during a test session without distracting them. The programming and recording equipment was located in another room some distance away so that relay clicks, etc. could not serve as conditioning cues. A clicker was presented via a loud-speaker in the isolation room throughout the daily test sessions to mask any extraneous noises from the corridors or other test rooms.

#### Preliminary Training

The animals were placed two at a time into individual primate chairs where they remained for the duration of the experiment. They were then given a 5-day period to become accustomed to this rather restrictive environment. On the sixth day avoidance conditioning was begun for each of the animals. A response lever was attached to the primate chair and a visual stimulus panel was located above and slightly to one side of the monkey's head. A flexible metal bracelet through which a shock stimulus could be delivered was placed upon the animal's right ankle. The other contact for shock was attached to the metal chair upon which the monkey was sitting. The two chairs were positioned so that the monkeys were back to back and a wooden screen was placed between the chairs so that the animals could not see each other.

The stimulus panel and lever were attached to the chair of one of the animals. A 15-minute period then elapsed to permit the monkey to adapt to these changes and to allow the experimenter to obtain a measure of the operant level for bar pressing. The

masking clicker was then turned on for the duration of the experimental session. Two minutes later a tape programmer was started which automatically presented the 20 daily trials. Initially a tape was programmed for a 30-minute session with trials randomly sequenced at 60, 90, or 120 seconds with an average intertrial interval of 90 seconds. However, during the conditioning of the first two subjects another tape was introduced spacing trials at an average of 2.25 minutes between the 20 trials for a 45-minute test session. This program, which was retained for the remainder of the experiments, presented trials at randomly sequenced intervals of 1.0, 1.5, 2.0, 2.5, and 3.0 minutes.

Each trial consisted of the presentation of the violet stimulus light for 6 seconds before the introduction of a shock delivered through the ankle bracelet and chair. A bar press terminated both the shock and the stimulus light. If the instrumental response occurred within 6 seconds of stimulus onset, the light was terminated and the shock was avoided.

When the conditioning session for the first animal was completed, the second monkey was then given a similar conditioning period. The first two animals (Numbers 120 and 132) received 10 such conditioning sessions for a total of 200 trials.

### *Communication of Affect*

Following the avoidance training in which each animal had learned to perform a bar pressing response to a violet stimulus light in order to avoid shock, the primate chairs were turned so that the animals faced each other and the wooden screen between them was removed. The ankle bracelets and metal chairs were wired in parallel so that the two monkeys would receive simultaneous shocks.

On the eleventh test day the stimulus panel was attached to Monkey 120's chair but no response bar was provided him. Number 132 was equipped with the response lever but did not have the stimulus panel. Test trials were presented on the same schedule as during the conditioning sessions and the shock unconditioned stimulus continued to be presented for failure to press the bar within 6 seconds of the stimulus light.

The task confronting the animals is obvious. The monkey with the light had no means of performing the instrumental response which would avoid the noxious shock stimulus. The second monkey could perform the response but had no stimulus to inform him when a response was appropriate. However, if the animal with the stimulus was able to communicate to his partner by means of expressive cues the information that the conditioned stimulus (CS) was being presented, the second monkey could then make the appropriate instrumental response which would enable both animals to avoid the shock.

Monkey Number 120 was given the stimulus panel and Number 132 the response bar for 22 consecutive test sessions. Following this, 4 days of individual reconditioning trials were given and, then, the roles of the two animals were reversed, Number 120

being given the bar and 132 the stimulus light, for 10 more testing sessions. On several sessions a wooden screen was placed between the animals which masked off all but facial cues from the monkey having the CS.

After this pair of animals had been tested in both directions, i.e., first 120 with the stimulus and then 132 with the stimulus, Monkey 132 was removed from his chair and replaced by Number 129. This animal had also been avoidance conditioned to visual stimuli in a primate chair several months prior to this experiment. However, he had not been exposed to the CS of a violet light. No preliminary training was administered to Monkey 129. After a 5-day period of adaptation to the chair, he was placed facing Number 120 in the communication of affect situation. The stimulus panel was mounted on 129's chair and the response lever was attached to 120's chair. Twenty trials per day were administered for 10 days. Then the wooden screen masking all but 129's head was introduced for two test sessions and, finally, a wooden barrier completely obstructing the animals' view of each other was placed between the chairs for two additional test sessions.

The positions of response lever and stimulus panel were again switched: 129 now having the response bar while 120 had the stimulus light. Ten more test sessions were administered to this combination with the partial screen on Days 6, 7, and 8 and the complete screen on Days 9 and 10.

### RESULTS

The initial adaptation to the confinement of the primate chairs and the establishment of a conditioned avoidance response in the first pair of animals were accomplished without incident. Since these animals had prior experience in a similar situation, learning was rapid and uneventful. Both animals performed avoidance responses on better than 95% of the trials on the last three conditioning days.

The communication of affect data had to be analyzed in terms of two response parameters. The first of these, of course, was the incidence of avoidances, i.e., trials on which a bar press occurred within 6 seconds of presentation of the CS to the stimulus animal. The second was the number of "spontaneous" or operant responses which occurred in the intervals between trials, i.e., in the absence of the conditioned stimulus. If the level of spontaneous responses were high, the probabilities of obtaining fortuitous or spurious avoidance responses would also be high. Therefore, in order to assess the level of authentic avoidances, the number of responses



TABLE 1

COOPERATIVE AVOIDANCE CONDITIONING:  
FIRST PAIR OF ANIMALS

Test day	Monkey with light	Monkey with bar	Number of avoidances	Estimated chance avoidances	<i>p</i>
1	120	132	3	9.8	<i>ns</i>
2	120	132	17	20	<i>ns</i>
3	120	132	17	14.1	<i>ns</i>
4	120	132	19	17.1	<i>ns</i>
5	120	132	18	18.6	<i>ns</i>
6	120	132	19	3.8	<.01
7	120	132	14	6.2	<i>ns</i>
8	120	132	16	10.8	<i>ns</i>
9	120	132	19	6.6	<.01
10	120	132	19	10.2	<.01
11	120	132	19	5.4	<.01
12	120	132	19	7.2	<.01
13	120	132	18	8.4	<.01
14	120	132	18	3.1	<.01
15	120	132	15	4.3	<.01
16	120	132	15	2.6	<.01 <sup>a</sup>
17	120	132	16	0.94	<.01 <sup>a</sup>
18	120	132	15	0.74	<.01 <sup>a</sup>
19	120	132	20	14.6	<i>ns</i> <sup>a, c</sup>
20	120	132	18	15.9	<i>ns</i> <sup>a</sup>
21	120	132	15	1.30	<.01 <sup>a, d</sup>
22	120	132	16	1.07	<.01 <sup>a, e</sup>
23	132	120	19	2.32	<.01
24	132	120	19	1.63	<.01
25	132	120	18	0.74	<.01
26	132	120	19	0.51	<.01
27	132	120	18	0.56	<.01 <sup>a</sup>
28	132	120	20	0.18	<.01 <sup>a</sup>
29	132	120	18	0.60	<.01 <sup>a</sup>
30	132	120	20	0.56	<.01 <sup>a</sup>
31	132	120	20	0.00	<.01 <sup>b</sup>
32	132	120	19	2.14	<.01 <sup>b</sup>

<sup>a</sup> Half screen between animals.<sup>b</sup> Full screen between animals.<sup>c</sup> Experimenter accidentally reduced bar tension to the point that chair shaking, etc. gave multiple bar presses.<sup>d</sup> Bar tension increased.<sup>e</sup> Discovery that misconnection of shock lead had prevented shock to Number 120 for previous 22 days.

occurring in the absence of the CS was divided by 430. This divisor was derived by dividing the total intertrial time (nonstimulus periods) into 6-second intervals. This calculation gave the number of spontaneous responses (SRs) expected per 6-second interval assuming a rectilinear distribution of such responses. The data indicated that such an assumption was, in general, justified although there was some tendency for SRs to diminish in frequency during the session and for SRs to occur with somewhat greater frequency immediately after a trial as a part of a chain of responses triggered by the CS. These departures from rectilinearity tended to bias the data against the hypothesis of communication of affect. There was never any suggestion of a temporal conditioning effect or a pattern of responses resembling fixed ratio or fixed interval operant conditioning.

The number of avoidance responses which would be expected on the basis of chance was derived by multiplying the number of SRs per 6-second interval by 20, the number of such intervals during which the CS was actually present. Since the number of spontaneous responses varied from day to day, a separate analysis was made for each testing session.

The data for the first pair of animals (120 and 132) is presented in Table 1. As the footnote indicates, the experimenters discovered after 22 testing days that the monkey with the stimulus (Number 120) had been improperly connected to the shock apparatus so that he had received no shock during the entire communication of affect test series. The *p* values were determined from the Mainland tables (Mainland, Herrera, & Sutcliffe, 1956) with the number of trials (20) as *N*. These data indicate that the animals performed a discriminated avoidance response in the social communication situation. The number of SRs was initially fairly high but diminished rapidly as the avoidance behavior became more discreet. The animal with the bar learned to attend very carefully to his partner with the CS and responded rapidly with bursts of bar presses when the stimulus monkey reacted to the CS. In this connection the behavior of Monkey 120 during those trials when he did not receive shock while Number 132 did was particularly interesting.

From observations made through the one-way vision windows it was impossible to detect the absence of shock from 120's behavior. He responded as 132 received the shock with a pronounced startle and leg withdrawal. It may be that this communication of affect (in the direction opposite from that being tested) served as reinforcement to maintain 120's affective response to the CS.

The data also indicate that visual cues alone were not the sole basis for solution of the mutual avoidance problem. The placing of either partial or complete screens between the animals did not eliminate successful avoidances. It was apparent from observation that auditory cues were also a factor in the communication process. The animal with



TABLE 2  
COOPERATIVE AVOIDANCE CONDITIONING:  
SECOND PAIR OF ANIMALS

Test day	Monkey with light	Monkey with bar	Number of avoidances	Estimated chance avoidances	<i>p</i>
1	129	120	15	20.0	ns
2	129	120	20	16.5	ns
3	129	120	17	20.0	ns
4	129	120	19	8.7	<.01
5	129	120	20	6.6	<.01
6	129	120	19	6.5	<.01
7	129	120	20	3.8	<.01
8	129	120	19	2.2	<.01
9	129	120	19	5.8	<.01
10	129	120	20	4.7	<.01
11	129	120	19	9.6	<.01 <sup>a</sup>
12	129	120	19	9.3	<.01 <sup>a</sup>
13	129	120	20	1.9	<.01 <sup>b</sup>
14	129	120	19	0.7	<.01 <sup>b</sup>
15	120	129	7	1.4	ns
16	120	129	20	7.9	<.01
17	120	129	20	2.0	<.01
18	120	129	18	0.24	<.01
19	120	129	20	1.3	<.01
20	120	129	20	7.4	<.01 <sup>a</sup>
21	120	129	20	3.8	<.01 <sup>a</sup>
22	120	129	20	0.8	<.01 <sup>a</sup>
23	120	129	18	11.8	<.01 <sup>b</sup>
24	120	129	20	8.6	<.01 <sup>b</sup>

<sup>a</sup> Half screen between animals.

<sup>b</sup> Full screen between animals.

the stimulus panel frequently squirmed and slapped at his chair when the stimulus was being presented. It was not possible to determine whether vocalization may also have occurred to the CS. That the results of the experiment were not attributable to spurious clicks, etc. from the apparatus itself was demonstrated by tilting the stimulus panel so that neither animal could see it and presenting a series of stimuli with the shock turned off. On the two occasions when this control procedure was run not a single bar press occurred while the CS was present.

The cooperative avoidance data from the second pair of animals, Numbers 120 and 129, are presented in Table 2. It may be recalled that Monkey 129 had no preliminary avoidance training to the stimulus employed in this investigation. As indicated in Table 2, 120 responded to the new situation by making large numbers of intertrial lever responses for the first 3 test days. The number of SRs was high enough so that merely by chance alone avoidance would be expected to occur on the majority of test trials. As Table 2 in-

dicates the two animals actually received only eight shocks during these first 3 days. Nevertheless, this relatively few pairings of the CS and UCS was sufficient to condition 129 and from Day 4 the discrimination between CS presentations and intertrial intervals became significant.

When this pair of animals was switched so that 129 had the response mechanism while 120 had the stimulus light, problem solution was very rapid. Monkey 129 which had never had the bar in this test situation attended very closely to the behaviors of Number 120 after the first test day and responded only rarely between trials (see Figure 1).

The insertion of screens between the animals in the second pairing seemed to have a more disrupting effect on the discrimination than in the initial pairing. This suggests that visual cues may have played a more important role in the communication of affect in the second pair of animals than in the first.

The data for all of the combinations of animals which were tested for mutual avoidance behavior revealed that when a discrimination had been established, the avoidance responses occurred as a short burst of bar presses. The first such response was, by definition, the avoidance since only a single response was required to terminate the CS and preclude the UCS. However, the number of responses occurring within the 6-second interval beginning with stimulus presentation

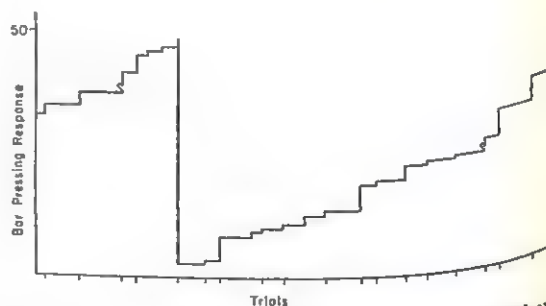


FIG. 1. Mutual avoidance conditioning. (Monkey 120 with stimulus light and 129 with bar. Trials 1 through 18 of Test Day 18 are shown. Diagonal jogs on Trials 3 and 16 indicate that avoidance was not successful and shock was delivered to both animals. Vertical pips on the abscissa indicate trials delivered at variable intervals by a tape programmer, from Trial 1 on the extreme left to Trial 18 on the extreme right. The 15 minute pretrial control period is not shown.)

was much greater than chance expectation. On several occasions the mean number of bar presses occurring within stimulus intervals was over 100 times greater than the frequency of such responses expected purely by chance.

#### DISCUSSION

The phenomenon of intraspecies communication has been prominently mentioned in the experimental literature on the infrahuman primates. Carpenter (1952) has stated, "Each known genus of primates has a repertoire of gestures which are employed consistently and which stimulate consistent reactions." The studies of cooperative problem solving (Crawford, 1941) and food sharing (Nissen & Crawford, 1936) in the chimpanzee indicate the importance and effectiveness of communication in that species. It was found that the chimpanzee could "solicit" food from a partner or direct its behavior in a cooperative serial problem solving task.

While the communication between primates has been recognized and described, there have been no systematic attempts to investigate the phenomenon experimentally. The present experiment represents an attempt to develop a methodology through which the communication processes may be more intensively studied. The data suggest that the "cooperative conditioning" technique is an exceptionally efficient and sensitive tool for the investigation of nonverbal communication. The method would appear to lend itself to the isolation and identification of the discrete expressive cues which convey affects between animals.

Some question arose before these animals were tested as to how the monkey with the response mechanism might be expected to behave in the social situation. While we hoped that a communication of affect would be utilized in the solution of the problem, there was a reasonable possibility that a second but less efficient solution would be learned. From the responding animals' viewpoint, an adequate avoidance could be achieved by setting up a rate of bar pressing which would insure that no more than 6 seconds elapsed between successive responses. In other words, the problem might have been structured by

the subject in the same fashion that monkeys and other species learn to avoid in the Sidman type of problem. There is no doubt that monkeys can learn to maintain a rather steady, high rate of responding in such a problem situation (Sidman, 1958). However, steady state responding at a high rate is a very costly adaptive mechanism from the subject's point of view if it must be maintained for any extended period of time. Porter, Brady, Conrad, Mason, Galambos, and Rioch (1958) demonstrated that severe, indeed, terminal gastrointestinal lesions develop in a remarkably short period of time in monkeys required to lever press every 20 seconds for a 6-hour test period. The Porter et al. study, incidentally, resembles in some ways the present experiment in that two animals were hooked together in a "yoked chair" setup. The experimental animal could prevent shock not only to himself but to his paired control by pressing the lever. The control animal, however, had no direct role in the avoidance behavior in that experiment while both animals in the present study played indispensable roles in problem solution.

It might also be pointed out that the present method can be viewed from several different viewpoints. While we were oriented primarily in terms of nonverbal communication of affect, it seems quite appropriate to consider these experiments as studies of cooperative behavior. By making some changes in initial training it would appear to be feasible to attempt to isolate some parameters of cooperation-competition in this experimental situation. Likewise, since the two animals are mutually dependent for a satisfactory solution to their joint problem, it might be possible to structure the experiment in such a way as to study the development of dependency relationships.

At the present time more intensive studies of the nonverbal communication of affect are being conducted. Improved control of the sensory modalities involved in the communication has been achieved so that the two animals may be isolated from each other, thus eliminating extraneous noises from twisting or shaking in the chairs plus vocal and olfactory cues. Further, joint reward conditioning,

and combined reward-avoidance schedules are being tested to determine the cues utilized in the communication of affects other than fear.

#### SUMMARY

Rhesus monkeys in primate chairs were conditioned to bar press within 6 seconds of presentation of a light in order to avoid electric shock. Following acquisition of this avoidance response two animals were placed facing each other and the bar was removed from the chair of one monkey and the stimulus light from the chair of the other. In order for either monkey to avoid shock a communication was necessary since neither animal had access to all elements of the problem. The results indicated that through nonverbal communication of affect an efficient mutual avoidance was performed. It was concluded that this paradigm is an exceptionally efficient and sensitive method for investigations of nonverbal communication.

#### REFERENCES

- CARPENTER, C. R. Social behavior of non-human primates. Structure et physiologie des sociétés animales. *Colloq. int. Cent. nat. Rech. scient.*, 1952, 34, 227-246.
- CRAWFORD, M. P. The cooperative solving by chimpanzees of problems requiring serial responses to color cues. *J. soc. Psychol.*, 1941, 13, 259-280.
- MAINLAND, D., HERRERA, T., & SUTCLIFFE, M. *Statistical tables for use with binomial samples*. New York: New York Univer. Press, 1956.
- MILLER, R. E., MURPHY, J. V., & MIRSKY, I. A. The modification of social dominance in a group of monkeys by interanimal conditioning. *J. comp. physiol. Psychol.*, 1955, 48, 392-396.
- MILLER, R. E., MURPHY, J. V., & MIRSKY, I. A. Nonverbal communication of affect. *J. clin. Psychol.*, 1959, 15, 155-158. (a)
- MILLER, R. E., MURPHY, J. V., & MIRSKY, I. A. The relevance of facial expression and posture as cues in the communication of affect between monkeys. *Arch. gen. Psychiat.*, 1959, 1, 480-488. (b)
- MIRSKY, I. A., MILLER, R. E., & MURPHY, J. V. The communication of affect in rhesus monkeys: I. An experimental method. *J. Amer. Psychoanal. Ass.*, 1958, 6, 433-441.
- MURPHY, J. V., MILLER, R. E., & MIRSKY, I. A. Inter-animal conditioning in the monkey. *J. comp. physiol. Psychol.*, 1955, 48, 211-214.
- MURPHY, J. V., & MILLER, R. E. The manipulation of dominance in monkeys with conditioned fear. *J. abnorm. soc. Psychol.*, 1956, 53, 244-248.
- NISSSEN, H. W., & CRAWFORD, M. P. A preliminary study of food-sharing behavior in young chimpanzees. *J. comp. Psychol.*, 1936, 22, 383-419.
- PORTER, R. W., BRADY, J. V., CONRAD, D., MASON, J. W., GALAMBOS, R., & RIOCH, D. Some experimental observations on gastrointestinal lesions in behaviorally conditioned monkeys. *Psychosom. Med.*, 1958, 20, 379-394.
- SIDMAN, M. By-products of aversive control. *J. exp. Anal. Behav.*, 1958, 1, 265-280.

(Received March 6, 1961)



## OPINIONS ABOUT MENTAL ILLNESS IN THE PERSONNEL OF TWO LARGE MENTAL HOSPITALS<sup>1</sup>

JACOB COHEN

*New York University*

AND

E. L. STRUENING

*Montrose Veterans Administration Hospital*

The past decade has witnessed several major shifts in the conception, care, and treatment of hospitalized mental patients. There has been a move toward "open" hospitals, milieu therapy, patient government, and patient work programs. This newer outlook is based on the general assumption that the well-being of mental patients is at least to some extent influenced by the social context. Derivations from this assumption include the more specific hypotheses that mental patients are sensitive to and influenced by the attitudinal atmosphere created by hospital employees, that the success of reintegrating former mental patients into society is affected by the attitudes of the general public toward mental illness, and that these attitudes play a role in determining the support of mental health programs by the general public as voters and taxpayers.

Despite the manifest importance of this area, there has been little systematic research directed toward the finding of relationships between attitudes toward the mentally ill and such variables as symptom reduction, successful rehabilitation of former patients, hospital discharge rates, etc. Research of this kind depends upon the adequate conception and objective measurement of attitudes toward mental illness and the mentally ill. The major

purpose of this investigation is to meet this need.

Previous work in this area is sparse. The pioneer work was the development of the Custodial Mental Illness Ideology (CMI) Scale by Gilbert and Levinson (1956). CMI scores were found to have correlates with hospital occupation, hospital treatment policy, and the California F Scale. The Cummings (1957) developed a Guttman scale of social distance from mental patients for use in their community studies. In both of these studies, a single (albeit different) pro-anti dimension of attitude was conceived, appropriate items were written, and were found to "scale" in the sense that internal consistency reliability (or reproducibility) was found to be adequate. But the fact that items can be organized on a single dimension does not mean that they are best so organized. A ready analogy comes to mind from the area of human ability. That one can identify a general intellectual factor in higher order domains (Thurstone, 1947) does not deny the existence or usefulness of group factors (verbal, numerical, space, etc.) in the understanding of intellectual functioning. The methodological base from which this investigation proceeds is that opinions about the mentally ill are potentially multidimensional, and the number and nature of these dimensions is an empirical issue, and not one to be assumed in advance; thus, the choice of multiple-factor analysis.

We conceive of attitudes as inferred variables which carry an affective or at least an adient-avoidant valence. Operationally, then, our responses reflect opinions, and the factors derived therefrom may<sup>2</sup> represent attitudes.

The purpose of the investigation was twofold:

<sup>2</sup> But not necessarily. As will be noted later, Factors A through D are understood to be attitudinal in this sense, while E is an opinion factor.

<sup>1</sup> This work was carried out at the Franklin D. Roosevelt Veterans Administration Hospital, Montrose, New York, as part of the Veterans Administration Psychiatric Evaluation Project, Richard L. Jenkins, Director. We are indebted to many Veterans Administration personnel for their cooperation, among them L. L. Rackow, Manager; S. G. Klebanoff, Chief Psychologist; and Gloria Fischer, now at the University of Oklahoma. We also wish to express our gratitude to the IBM Watson Scientific Computing Laboratory, New York, for training us in their facilities. Finally, we wish to express a continuing sense of indebtedness to Catherine S. Henderson, the project secretary, and Ethel Haas, our research assistant.

1. To identify and develop measures of the salient dimensions underlying opinions about severe mental illness among hospital personnel.

2. To explore the construct validity of these measures by relating them to demographic characteristics of the respondents—occupation, education, age, and sex.

## METHOD

*Items.* A pool of approximately 200 opinion items referring to the cause, description, treatment, and prognosis of severe mental illness was prepared. These items were made up of quotations from case conferences and casual conversations, and paraphrases of ideas which are current in the mental hospital. This group of items was reviewed by a group of hospital experienced research workers,<sup>3</sup> and items were edited, balanced with regard to pro and anti content, and overlapping ones discarded until 55 remained. These were supplemented by items (in some cases revised) from the Custodial Mental Illness Ideology (CMI) Scale (Gilbert & Levinson, 1956), the California F Scale (Struening, 1957), and Nunnally's (1957) work on popular conceptions of mental health to form a 70-item set. All items were presented in Likert format with provision following each for a checked response on a six-point agreement continuum.

*Subjects.* The phase of the investigation described here was carried out in two large Veterans Administration neuropsychiatric hospitals, one in the Northeast (Hospital I) and another in the Midwest (Hospital II) and utilized as subjects large samples of the personnel in these two hospitals. Later tables (7 and 8) give the distributions of these samples by occupation, education, age, and sex. There were in all 541 usable questionnaires from Hospital I and 653 from Hospital II, in each case broadly representative of the different levels and functions of personnel whose work brought them into frequent contact with the patients. In each hospital, some two-thirds to three-quarters of the target population were obtained as subjects.

*Administration.* Most of the subjects in both hospitals were group tested so as to guarantee anonymity. Where this was not possible (e.g., night nurses and aides who could not leave their posts), envelopes in which the completed questionnaires could be anonymously returned were provided. Instructions stressed anonymity and the fact that the issues were matters of opinion about which even professionals differed, so that there were no right or wrong answers.

*Data analysis.* The method of analysis described below was applied to each hospital separately. The two sets of data were analyzed successively, and in

<sup>3</sup> Seymour Slovik, Leonard Solomon, Herbert Spohn, Herbert Turkel, and Harold Wilensky participated in this phase of the work, for which we express our indebtedness.

exactly the same way. First, the 70 × 70 matrix of product-moment coefficients of correlation between individual items was computed. Since the available IBM 650 program for centroid factor extraction is limited to 40 × 40 matrices, the following strategem was employed in analyzing the data for Hospital I:

1. The five items which yielded no more than one correlation numerically greater than .20 were dropped from further consideration.

2. Fifteen items whose correlations varied greatly and would therefore be the nucleus of maximally independent item clusters (Tryon, 1958, p. 12) were selected as marker variables.

3. The remaining 50 items were randomly divided in half and to each half the marker items were attached, yielding two 40 × 40 matrices sharing a group of 15 marker items.

4. The two matrices were each subjected to a standard centroid factor analysis. The first five centroids<sup>4</sup> were rotated to an orthogonal, simple-structure type of solution by means of the quartimax analytic method (Neuhauser & Wrigley, 1954).

5. The two rotated factor matrices were then compared by reference to the 15 marker item loadings. They proved remarkably similar, and no difficulty was encountered in matching the five rotated factors from the two analyses.

6. The factor matrices for the two analyses were then recombined. Since each marker item had two sets of loadings (one from each rotation in which it figured), a single loading on each factor was obtained by determining the root mean square of the paired loadings (i.e., the square root of the average of the two squared loadings).

Exactly the same method was used to analyze the Hospital II data. In Hospital II, 6 items were dropped completely on grounds of low correlation and 16 were used as marker items, resulting again in two 40 × 40 matrices of intercorrelations sharing a group of marker items which were analyzed as above and recombined into a single factor matrix.

## RESULTS

### Factors

As was the case in matching the factors over the two 40 × 40 matrices within each hospital, no difficulty was encountered in matching the five rotated factors in the two hospitals. As evidence of the factor similarity between hospitals, for each factor separately the rotated factor loadings for the two hospitals were correlated over the 62 items which

<sup>4</sup> There are undoubtedly more than five common factors in each of the matrices, but starting with the sixth centroid, factor loadings do not exceed .30 and exceed .20 for only two or three items. Thus, the 5 factors extracted are deemed the most salient of a set of perhaps 15 (judging from the latent roots) common factors.



TABLE 1  
FACTOR A—AUTHORITARIANISM

Loading		Item
Hospital I	Hospital II	
76	61	68. There is hardly anything lower than a person who does not feel a great love, gratitude, and respect for his parents.
72	53	65. Obedience and respect for authority are the most important virtues children should learn.
61	55	11. When a person has a problem or worry, it is best not to think about it, but keep busy with more pleasant things.
61	49	34. A heart patient has just one thing wrong with him, while a mentally ill person is completely different from other patients.
58	51	16. All patients in mental hospitals should be prevented from having children by a painless operation.
56	54	26. There is something about mental patients that makes it easy to tell them from normal people.
60	47	29. People with mental illness should never be treated in the same hospital as people with physical illness.
59	49	14. Mental illness is usually caused by some disease of the nervous system.
47	58	27. If people would talk less and work more, everybody would be better off.
52	49	60. Every person should make a strong attempt to raise his social position.
42	55	22. It is easy to recognize someone who once had a serious mental illness.
58	36	12. Nervous breakdowns usually result when people work too hard.
53	42	8. People who are mentally ill let their emotions control them; normal people think things out.
49	46	52. Although patients discharged from mental hospitals may seem all right, they should not be allowed to marry.
58	32	2. One of the main causes of mental illness is a lack of moral strength or will power.
43	50	59. Every mental hospital should be surrounded by a high fence and guards.
52	40	6. People would not become mentally ill if they avoided bad thoughts.
53	35	70. Every person should have complete faith in some supernatural power whose decisions he obeys without question.
51	38	31. A person who has bad manners, habits, and breeding can hardly expect to get along with decent people.
45	41	50. The best way to handle patients in mental hospitals is to keep them behind locked doors.
54	23	39. Although some mental patients seem all right, it is dangerous to forget for a moment that they are mentally ill.
45	39	66. College professors are more likely to become mentally ill than are business men.
39	40	24. Regardless of how you look at it, patients with severe mental illness are no longer really human.
-72	06	13. The patients of a mental hospital should have something to say about the way the hospital is run. <sup>a</sup>

<sup>a</sup> Not in root mean square order because of inconsistency in loading.

appeared in both analyses. The resulting Pearson correlations between hospitals are: A, .86; B, .73; C, .38; D, .60; E, .77. When the "unadjusted correlation" (Burt, 1941, p. 343) is used as an index of similarity, the resulting values are: A, .92; B, .62; C, .51; D, .61; E, .81. Using either criterion the results, with the possible exception of Factor C, are deemed satisfactory.

The similarity of the factors between hospitals can be further judged by reference to Tables 1 through 5, where the factor loadings of both hospitals for the most highly loaded items are presented. In each table, the items

are presented in decreasing order of their root mean square loadings over the two hospitals (which are, however, not given) down to the level of .40 in Table 1 and about .30 in Tables 2-5.<sup>5</sup> Decimal points are omitted. Positive loadings indicate agreement, negative loadings disagreement.

<sup>5</sup> The only noteworthy inconsistency in loading occurs for Item 13 which loads Factor A heavily (-.72) in Hospital I and Factor C even more heavily (.88) in Hospital II (Tables 1 and 3). Note that this is not a logical inconsistency; the item fits both interpretations.



TABLE 2  
FACTOR B—BENEVOLENCE

Loading		Item
Hospital I	Hospital II	
44	44	28. Even though patients in mental hospitals behave in funny ways, it is wrong to laugh about them.
-42	-41	54. There is little that can be done for patients in a mental hospital except to see that they are comfortable and well fed.
51	35	41. Anyone who tries hard to better himself deserves the respect of others.
30	41	37. Patients in mental hospitals are in many ways like children.
-44	-26	35. To become a patient in a mental hospital is to become a failure in life.
-44	-07	42. Our mental hospitals seem more like prisons than like places where mentally ill people can be cared for.
26	37	4. Although they usually aren't aware of it, many people become mentally ill to avoid the difficult problems of everyday life.
12	43	38. More tax money should be spent in the care and treatment of people with severe mental illness.
27	34	39. Although some mental patients seem all right, it is dangerous to forget for a moment that they are mentally ill.
30	29	60. Every person should make a strong attempt to raise his social position.

*Factor A—Authoritarianism.* The conception of mental patients projected by this factor is one which stresses their difference from and inferiority to normal people (Items 34, 16, 26, 29, 22, 8, 52, 39, 24, and 13). Several items present popular (and contradictory) ideas about the causality of mental illness (Items 14, 12, 2, 6, and 66).

This view of the mental patient exists in a context which results in the five items (Items 27, 31, 65, 68, and 70) taken from the California F Scale (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950) having high loadings on this factor; indeed, the two items giving the largest loadings on Factor A are from the F Scale. These reflect the characteristic submission to authority (Items 65, 68, and 70) and "anti-intraceptiveness" (Items 11, 27, 6, and 66) of the authoritarian. In fact, Items 6 and 66 indict thinking (bad or too much) as playing an etiological role in mental illness. The handling of the hospitalized mentally ill advocated here, namely, high fence, guards, locked doors (Items 59 and 50) bears the coercive authoritarian stamp.

We have named the common factor defined by the above elements Authoritarianism. It presents a gestalt made up of authoritarian submission and anti-intraception with a view of the mentally ill as a class inferior to normals and requiring coercive handling. A most interesting possibility suggests itself that for

the authoritarian personality within the mental hospital, the mentally ill may function as a negatively stereotyped outgroup in much the same way as do racial, religious, or political minority groups in the larger society. Indeed, Factor A is essentially identical with what the F Scale measures. This is demonstrated by the fact that the correlation between Factor A scores based solely on items having mental illness content (see below) with scores obtained by summing the 6 F Scale items is .86, and exceeds unity when corrected for attenuation.<sup>6</sup>

Factor A is a dominant factor which accounts for an average of 47% of the common variance in the two hospitals. This incidentally suggests that it is also essentially what is measured by the CMI (Gilbert & Levinson, 1956), since the total score obtained by adding together such items will be richly saturated in the largest common factor running through the item set, at the expense of less extensive common factors. The substantial correlation between CMI and F was demonstrated by Gilbert and Levinson (1956).

*Factor B—Benevolence.* Factors B (Table

<sup>6</sup> The role of acquiescence set in Factor A was not explicitly studied. However, its importance cannot be great in the light of the fact that in Hospital I, Item 13 gives rise to a negative loading of .72, i.e., those otherwise high on Factor A disagree with high consistency with this item.

2) and C (Table 3) are both "promental patient," but they are so from rather different perspectives, as evidenced by their near zero shared variance (Table 6). The positive pole of B represents a benevolence toward patients which arises from a moral point of view, a sort of Christian kindness toward unfortunates. Mental patients are seen *not* as failures in life (Item 35), but rather are like children (Item 37), and it is wrong to laugh about them (Item 28). Still, it is dangerous to forget for a moment that they are mentally ill (Item 39), a point of view which is also part of Factor A. They are looked upon as an obligation of society (Item 38), and more than mere custodial care should be offered them (Item 54). Still, mental hospitals are *not* like prisons (Item 42), in contrast with the Factor C view. Finally, Items 41 and 60 project the traditional value of self-improvement, which in this context suggests its advocacy to mental patients (but note that Item 60 also loaded Factor A). Factor B accounts for an average of 15% of the variance shared by the items.

In our earlier reports on this investigation (Cohen & Struening, 1959, 1960), we qualified "benevolence" by the adjective "unsophisticated," which was not meant pejora-

tively, but intended to distinguish this factor from Factor C. Others close to our work have responded to other aspects of this factor and have suggested such names as Parental Benevolence, Moral Benevolence, and Humanistic Benevolence. All these qualifiers can be supported by the item loadings. We have settled on the "common factor" of these interpretations, Benevolence, as the verbal tag to assign this factor. What is intended is a kindly, paternalistic view towards patients whose origin is in religion and humanism rather than a scientific or professional dogma. It is encouraging and nurturant, but still acknowledges some fear of patients. Further support of this interpretation will be offered in the later discussion of its demographic correlates.

*Factor C—Mental Hygiene Ideology.* Factor C involves an orientation toward mental patients which is also positive, but embodies the tenets of the creed of modern mental health professionals (Table 3). The items here are more factually descriptive of the mentally ill; e.g., they are willing to work (Item 23), capable of skilled labor (Item 53), many would remain with unlocked doors (Item 55), ex-patients could be trusted as baby sitters (Item 61), there is much mental

TABLE 3  
FACTOR C—MENTAL HYGIENE IDEOLOGY

Loading		Item
Hospital I	Hospital II	
48	25	38. More tax money should be spent in the care and treatment of people with severe mental illness.
45	28	44. If our hospitals had enough well trained doctors, nurses, and aides, many of the patients would get well enough to live outside the hospital.
a	37	23. Most mental patients are willing to work.
26	37	61. Most women who were once patients in a mental hospital could be trusted as baby sitters.
40	19	53. Many mental patients are capable of skilled labor, even though in some ways they are very disturbed mentally.
39	20	25. Many people who have never been patients in a mental hospital are more mentally ill than many hospitalized mental patients.
31	a	69. The death penalty is inhuman and should be abolished.
31	31	55. Many mental patients would remain in the hospital until they were well, even if the doors were unlocked.
10	41	42. Our mental hospitals seem more like prisons than like places where mentally ill people can be cared for.
37	16	21. Mental illness is an illness like any other.
14	88	13. The patients of a mental hospital should have something to say about the way the hospital is run. <sup>b</sup>

a Item omitted from the analysis.

b Not in root mean square order because of inconsistency in loading.

illness outside of hospitals (Item 25). Implicit in this conception is the idea that mental patients are much like normal people, differing from them perhaps in degree, but not in kind, in sharp contrast with the Factor A orientation. This view is partially summarized in "Mental illness is an illness like any other" (Item 21). The efficacy of treatment is strongly believed in (Item 44), as is the assumption by society of its obligations to the mentally ill (Item 38), the latter shared with Factor B. In contrast with Factor B, the item likening mental hospitals to prisons (Item 42) is endorsed here. Finally, opposition to the death penalty (Item 69) and advocacy of "hospital democracy" (Item 13) round out the picture. Factor C accounts for an average of 14% of the shared variance.

Because of the concordance of the elements of Factor C with the tenets of the mental hygiene movement, this factor was named Mental Hygiene Ideology. This interpretation is further supported by its correlation with occupation and education (see below).

*Factor D—Social Restrictiveness.* Factor D, which accounts for an average over the two hospitals of 14% of the shared variance, emphasizes the desire to restrict mental patients both during and after hospitalization

for the protection of society, particularly the family unit (see Table 4). Thus, they should not be allowed to marry after hospitalization (Item 52), should be easily divorced upon hospitalization (Item 48); indeed a woman "would be foolish" to marry an ex-mental patient (Item 40). Their parental rights should also be restricted by forbidding their small children from visiting them (Item 20) and by sterilization (Item 16). Posthospital employment as baby sitters should be closed to them (Item 61). Mental patients should be denied the right to vote (Item 57). All these items share the belief that mental illness is a threat to society which must be met by some restriction in social functioning both during and following hospitalization; Factor D was accordingly named Social Restrictiveness. The other items loading this factor account for the basis of the restrictive orientation. Thus, mental patients are seen as socially deficient; they do not care how they look (Item 64) and do not make wholesome friendships (Item 51). Furthermore, the outlook for their future is hopeless: "There is little that can be done" for them (Item 54)—they "Will never be their old selves again" (Item 43).

One notes a certain similarity in outlook between Factor D and Factor A, and, in fact,

TABLE 4  
FACTOR D—SOCIAL RESTRICTIVENESS

Loading		Item
Hospital I	Hospital II	
49	52	40. A woman would be foolish to marry a man who has had a severe mental illness, even though he seems fully recovered.
45	49	52. Although patients discharged from mental hospitals may seem all right, they should not be allowed to marry.
36	55	43. People who have been patients in a mental hospital will never be their old selves again.
13	50	54. There is little that can be done for patients in a mental hospital except to see that they are comfortable and well fed.
35	33	48. The law should allow a woman to divorce her husband as soon as he has been confined in a mental hospital with a severe mental illness.
-17	-42	61. Most women who were once patients in a mental hospital could be trusted as baby sitters.
38	22	20. The small children of patients in mental hospitals should not be allowed to visit them.
18	37	64. Most patients in mental hospitals don't care how they look.
16	37	16. All patients in mental hospitals should be prevented from having children by a painless operation.
-27	-29	51. Many patients in mental hospitals make wholesome friendships with other patients.
27	a	57. Anyone who is in a hospital for a mental illness should not be allowed to vote.

a Item omitted from analysis.



TABLE 5  
FACTOR E—INTERPERSONAL ETIOLOGY.

Loading		Item
Hospital I	Hospital II	
56	48	3. Mental patients come from homes where the parents took little interest in their children.
53	47	5. The mental illness of many people is caused by the separation or divorce of their parents during childhood.
52	47	1. If parents loved their children more, there would be less mental illness.
40	30	9. If the children of mentally ill parents were raised by normal parents, they would probably not become mentally ill.
35	23	33. If the children of normal parents were raised by mentally ill parents, they would probably become mentally ill.
32	25	4. Although they usually aren't aware of it, many people become mentally ill to avoid the difficult problems of everyday life.
27	28	67. People who are successful in their work seldom become mentally ill.

inspection of the graphed factor plots indicates that were the rotation oblique, there would be some correlation between A and D (see Table 6 and discussion below). These two factors are nevertheless distinct with the emphasis of restrictiveness to protect the family the distinguishing feature.

*Factor E—Interpersonal Etiology.* Factor E is highly specific in its reference and highly consistent between the two hospitals (see Table 5). Its positive pole reflects quite strongly a belief that mental illness arises from interpersonal experience, particularly deprivation of parental love and attention during childhood (Items 3, 5, and 1), or more generally the mental health of those *in loco parentis* (Items 9 and 33). Somewhat less central is a belief that abnormal behavior is motivated; e.g., mental illness is an avoidance of problems (Item 4), successful people seldom become mentally ill (Item 67). Accordingly, Factor E has been named Interpersonal Etiology. It accounts for 10% of the shared variance in each hospital.

#### Factor Scores

To obtain measures of each factor, items which loaded each factor most highly and which had mental illness content were composed by assigning integral values from one to six to the six alternatives from strongly disagree to strongly agree for the positively loaded items, and in reverse order for the negatively loaded items. Item scores were not standardized, since their standard deviations did not vary greatly, nor were weighted as a

function of their factor loadings, since it has been demonstrated that this refinement has little advantage over the simpler unit weights (Trites & Sells, 1955). Each item used for factor scoring was assigned to only one factor, in order to keep the factor scores experimentally independent. Constants were added, as necessary, to avoid negative scores.

The psychometric characteristics of the factor scores are given in Table 6. The reliability coefficients given are of the internal consistency type, equivalent to those obtained by the generalized Kuder-Richardson Formula 20 (Tryon, 1957), and are to be understood as giving an estimate of the correlation one would obtain from composites of the same number of items drawn randomly from the same item domain (Tryon, 1957). They are deemed quite satisfactory for our research purposes, possibly excepting Factor D. However, in the light of Factor D's higher factor validity coefficients, and significant demographic correlates (see below), its reliability is apparently adequate.

The factor validity coefficients were computed by means of Thomson's (1951, pp. 197-199) pooling square. The resulting value represents the correlation coefficient between the sum of a set of item scores and the factor they share in common. The validity coefficients are also quite satisfactory for the purposes of group comparisons.

Although the abstract factors defined by the quartimax rotation are mutually independent, the factor scores show some small correlations (Table 6). In both hospitals one

TABLE 6  
RELIABILITIES, FACTOR VALIDITIES, AND INTERCORRELATIONS  
OF THE OMI FACTOR SCORES IN HOSPITALS I AND II

Scale	n <sup>a</sup>	Reliability		Validity		Intercorrelations <sup>b</sup>				
		I	II	I	II	A	B	C	D	E
A	13	82	76	89	73		-19	-39	22	16
B	11	49	62	67	76	-20		11	-15	03
C	11	60	61	56	65	-26	28		-28	25
D	9	21	23	43	67	20	-12	-30		-02
E	7	60	59	78	66	08	14	24	05	

Note.—Decimal points omitted.  
<sup>a</sup> Number of items.  
<sup>b</sup> Hospital I above diagonal; Hospital II below.

finds a weak link between Authoritarianism (Factor A) and Social Restrictiveness (Factor D), another between Benevolence (Factor B) and Mental Hygiene Ideology (Factor C), and negative correlations across these two pairs. Interpersonal Etiology (Factor E) has small correlations with Mental Hygiene Ideology (Factor C), but, perhaps surprisingly, is uncorrelated with Authoritarianism (Factor A) and Social Restrictiveness (Factor D). It must be stressed that these relationships are quite small; of the 20 correlations in the two hospitals, only one exceeds .30.<sup>7</sup> One must be prepared to find all patterns of high and low scores on the five factors; no factor score predicts another to any material degree. Respondents low on Authoritarianism are no more likely to accept Interpersonal Etiology than those high on this factor. Respondents high on Mental Hygiene Ideology are only slightly more likely to reject Social Restrictiveness than those low on this factor.

Demographic Correlates

To study the relationships between the OMI factors and occupation, education, age, and sex, each hospital sample was broken down into subgroups for each demographic variable and simple analyses of variance performed on each OMI factor score. In addition to determining the significance of the departure from the overall null hypothesis, each analysis was extended to yield an eta coefficient of nonlinear correlation to make pos-

sible an assessment of the *degree* of association between the demographic variable and factor score involved.<sup>8</sup> Where applicable, the form of the regression of factor on demographic variable was determined.

*Occupation.* Table 7 presents the mean factor scores for 10 occupational groups who have daily patient contacts. Physical Medicine and Rehabilitation and Special Services are charged with carrying on "activity" therapies with patients. The mental health professional groups include residents and trainees. The Physicians groups are made up predominantly of nonpsychiatric physicians, but include dentists and chaplains. The occupational groups are significantly ( $p < .01$ ) differentiated on all factors in both hospitals, excepting only Social Restrictiveness in Hospital II. As can be seen from Table 7, the subgroup differences are particularly marked on Authoritarianism (Factor A), where the eta values .65 and .50 indicate (through their squares) that 42% and 25% of the score variance is associated with occupation subgroup membership. The correlations with occupation are also substantial for Benevolence (Factor B) and Mental Hygiene Ideology (Factor C), but noticeably smaller for the remaining two factor scores.

The sheer volume of data in Table 7 precludes detailed discussion, but certain highlights may be noted:

1. On Authoritarianism (Factor A), as might be expected, psychologists, psychiatrists, and social workers have low means, while those of aides and kitchen personnel

<sup>7</sup> A reviewer points out that these correlations are attenuated by the measurement error variance in the factor scores. This must be granted. Still, when they are corrected for attenuation, they reflect less than 50% shared variance in all but two or three instances.

<sup>8</sup> In the case of sex, the analogous *t* ratios and point biserial correlation coefficients were found.

TABLE 7  
MEANS AND ANALYSIS OF VARIANCE RESULTS OF OMI FACTOR SCORES  
BY OCCUPATIONAL GROUPS IN HOSPITALS I AND II

Occupation	N <sub>I</sub>	N <sub>II</sub>	Factor A		Factor B		Factor C		Factor D		Factor E	
			I	II	I	II	I	II	I	II	I	II
Clerical	19	33	18.8	19.3	44.3	46.3	34.1	36.4	20.3	19.3	19.3	21.7
Physical Medicine and Rehabilitation	27	25	19.6	23.8	44.3	44.1	35.8	38.5	20.7	18.7	19.7	19.3
Nurses	35	53	16.5	21.2	45.5	45.1	36.5	34.1	20.5	20.8	20.1	19.6
Aides	254	317	28.4	27.8	43.7	43.8	33.0	32.5	19.1	20.3	18.6	19.1
Psychiatrists	13	5	13.4	14.8	43.8	40.8	38.7	35.6	20.8	18.8	21.2	24.2
Social Workers	13	6	13.4	16.5	40.9	41.0	37.9	41.5	19.2	19.7	16.9	21.0
Physicians	11	10	19.5	22.2	42.5	43.2	34.1	37.6	23.4	23.5	20.4	21.4
Psychologists	36	19	11.9	15.0	37.4	38.8	43.6	40.1	16.5	18.2	21.6	23.1
Special Services	18	14	14.9	17.5	45.8	44.6	36.9	37.1	19.6	19.0	16.9	19.2
Kitchen Workers	94	120	31.1	31.3	42.0	41.8	33.0	32.9	19.9	19.6	20.0	18.2
(Total Sample)	(541)	(653)	(24.8)	(26.4)	(43.1)	(43.5)	(34.6)	(33.6)	(19.4)	(20.0)	(19.2)	(19.2)
<i>F</i> <sub>a</sub>			41.1	22.0	7.9	5.7	14.3	9.1	3.5	1.8	2.6	3.7
<i>p</i>			.01	.01	.01	.01	.01	.01	.01	ns	.01	.01
<i>eta</i> <sup>2</sup>			.65	.50	.35	.28	.45	.35	.24	.16	.21	.23

\* For Hospital I *df* = 9/510; for Hospital II *df* = 9/592.

are high. Gilbert and Levinson (1956) found exactly parallel occupational differences on the CMI, which again attests to the similarity between their scale and Factor A.

2. Psychologists again occupy the low extreme on Benevolence (Factor B), while the high end is taken by Special Service personnel, nurses, and ward clerical personnel. This latter finding does not imply that psychologists are malevolent—it is rather the moralistic-paternalistic perspective of Factor B which they reject.

3. Aides and kitchen workers have the lowest means on Mental Hygiene Ideology (Factor C), and psychologists, social workers, and Hospital I psychiatrists have the highest means, a state of affairs inversely

paralleling that of Authoritarianism (Factor A).

4. While Social Restrictiveness (Factor D) does not spread the groups apart strongly, it is striking to find in each hospital that the physician subgroup is by far the highest, while psychologists are the lowest.

5. Finally, and quite predictably, psychologists and psychiatrists most strongly accept Interpersonal Etiology while aides and kitchen workers tend to be low.

Differences in orientation towards the mentally ill among the various occupational groups are striking. Typically, the mental health professionals are at opposite poles from the aides, who provide the quantitative bulk of the "normal" social atmosphere for

TABLE 8  
MEANS AND ANALYSIS OF VARIANCE RESULTS OF OMI FACTOR SCORES  
BY YEARS OF EDUCATION IN HOSPITALS I AND II

Years of education	N <sub>I</sub>	N <sub>II</sub>	Factor A		Factor B		Factor C		Factor D		Factor E	
			I	II	I	II	I	II	I	II	I	II
- 8												
9-11	46	120	31.7	31.5	42.4	42.6	33.2	31.8	19.4	19.6	20.2	18.4
12	129	161	29.8	29.8	42.7	43.1	33.2	32.9	19.2	19.9	18.7	18.8
13-15	144	177	27.9	25.5	44.0	43.9	33.0	33.3	19.4	20.3	19.4	18.9
16-18	60	86	22.4	24.2	45.8	45.5	33.9	33.7	19.7	20.4	18.7	20.1
19-	79	62	15.8	19.0	43.2	44.2	36.8	37.8	20.7	19.8	18.8	20.6
<i>F</i> <sub>a</sub>	58	27	13.1	15.3	40.6	41.0	40.9	38.6	18.2	19.0	21.1	23.3
<i>p</i>			71.8	44.5	6.7	5.5	19.2	13.1	2.0	0.8	2.4	6.5
<i>eta</i> <sup>2</sup>			.01	.01	.01	.01	.01	.01	ns	ns	.05	.01
			.64	.51	.25	.20	.40	.31	.14	.08	.15	.22

\* For Hospital I *df* = 5/511; for Hospital II *df* = 5/628.



patients, the nurses falling in between (but note that Factors B and D depart from this pattern). The existence of such diversity of view attests to a host of problems in communication in the occupational hierarchy and consequently in patient care.

*Education.* This variable is related to occupation and parallels it for the OMI factors (see Table 8). Again the two hospitals are very similar. Since years of education is an ordered variable, it is possible to consider the shape of the regression of factor score on education, as well as the significance and degree of relationship.

1. Authoritarianism (Factor A) shows a sharp negative linear relationship with education of the order of .6 and .5 for the two hospitals, quite similar to the degree of relationship provided by the highly correlated occupation grouping. This is quite in keeping with the repeated findings of negative correlation of education and F Scale score, which, as already noted above, is considered to be essentially the same factor as A.

2. Benevolence (Factor B) has a weaker but significant inverted U shaped relationship with education in both hospitals, the peak coming in the "some college" (13-15) group. It drops off at higher levels, again presumably because of increased rejection of the moralistic basis for the benevolence rather than the benevolence itself. The low level of Factor B at elementary and high school levels is puzzling. Perhaps the Authoritarianism at these educational levels precludes Benevolence, just as Mental Hygiene Ideology at high educational levels precludes it.

3. Education is somewhat more strongly related to Mental Hygiene Ideology (Factor C), accounting for about 10-15% of its variance. The regression is such that the means remain constant from elementary through some college education and then rise sharply with graduation and postgraduate training. This finding in conjunction with the related finding on occupation supports the interpretation of Factor C as the mental health professional's creed.

4. Social Restrictiveness (Factor D) fails to relate significantly to education and was slightly (Hospital I) and nonsignificantly (Hospital II) related to occupation. Although this is psychometrically the weakest of the

factor scores, it nevertheless has enough factor validity to disclose an important relationship where it exists, given the large samples studied here (see age and sex below). The conclusion that Factor D does not relate to education, therefore, seems justified and further helps distinguish it from Authoritarianism, which is strongly related to education (and occupation).

5. Although significant, the relationship between years of education and Interpersonal Etiology (Factor E) is weak. Insofar as it can be discerned, in Hospital II the regression is positive with a sharp rise at the postgraduate level; in Hospital I the latter rise also appears, but it zigzags anomalously at lower levels.

*Age.* The OMI factor scores do not relate as strongly to age as they do to occupation and education.<sup>9</sup> Neither Benevolence (Factor B), nor Mental Hygiene Ideology (Factor C), nor Interpersonal Etiology (Factor E) are significantly related to age in either hospital, and the remaining factors are not strongly related.

1. Age accounts for about 5% of the variance on Authoritarianism in both hospitals, but does so in an unanticipated U shaped regression whose trough comes in the thirties. The conservatism of middle age is an often observed phenomenon, but no reason comes to mind for the Authoritarianism (as high or higher) of the teens and twenties. This may merely be a peculiarity of the age-education-occupation structure of these two mental hospitals which deserves no general inferences. In any case, the relationship is a weak one.

2. The relationship of age to Social Restrictiveness is even weaker and takes the form of a slow increase in score with age.

*Sex.* Like age, sex is only weakly related to the factor scores.<sup>9</sup> The sex differences on Authoritarianism (Factor A), Mental Hygiene Ideology (Factor C), and Interpersonal Etiology (Factor E) are not significant. The

<sup>9</sup> Tables giving these detailed findings have been deposited with the American Documentation Institute. Order Document No. 7057 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

These tables are also available from the authors.

others are significant in both hospitals, but of little consequence, in no instance accounting for as much as 3% of the variance.

1. Women show somewhat higher Benevolence (Factor B) scores on the average than men, a fact that accords with their cultural role. (It should be recalled, however, that nurses were the highest occupation subgroup on this factor and account for a quarter to a third of the sample of women.)

2. Women also show somewhat higher Social Restrictiveness (Factor D) than men, but this may be an artifact. Several of the items on this scale either explicitly (Items 40 and 48) or implicitly (Item 52) are couched in terms of what women should do in marital relationships with male mental patients. Were the shoe on the other foot, this difference might disappear or be reversed.

#### DISCUSSION

The Gilbert-Levinson (1956) CMI can be interpreted in the light of the results of the factor analysis. Their initial conception of custodialism-humanism has our Factors A and D at the custodialism pole, C and by implication E at the humanism pole, and B scattered all along it. The tendency for Authoritarianism (Factor A) to dominate collections of items such as they and we used taken with their Likert method of procedure resulted in their final CMI scale being largely a measure of Factor A. This leaves the other dimensions we have uncovered unaccounted for in their scheme.

The importance of this omission cannot, at present, be fully assessed, since the role of the other dimensions with regard to the efficacy of patient care has not yet been studied. We speculate, however, that more than Authoritarianism is important in regard to hospital, family, and community atmosphere and its relation to the well-being of mental patients. For example, we think that Benevolence (Factor B) is an important quality in psychiatric aides and nurses, who provide the bulk of the patients' contact with "normals," and represent the hospital to them. We believe that present educational programs for psychiatric aides seek to inculcate in them the ideology of the mental hygiene movement (Factor C), which is at least foreign,

quite possibly anxiety provoking, and in any case not very effective. This probably accounts for the failure of the Cummings' (1957) community education effort; the conception "Mental patients are different from you only in degree," a Factor C idea, proved flatly unacceptable to the community. Perhaps the message, "Mental patients are poor unfortunates whom we should help out of simple human kindness," a Factor B formulation, might have proven more effective, despite its condescending sound in the ears of professionals.

Such speculation as the above, although interesting and perhaps even exciting, is not meant to replace the empirical work that needs to be done, but merely to provide a basis for generating hypotheses for such work.

The substantial differences in factor scores found as one goes up the occupational-educational hierarchy of the two mental hospitals studied, particularly in the light of the consistency found between these widely separated hospitals, merits thoughtful attention. At least some of the friction found between professional groups in hospitals and some of the failures in communication between those who give orders and those who carry them out is in manifestation of widely separate views of the nature and progress of mental illness held by different occupational groups. The diagnosis of this problem (as in psychiatry generally) leads to no immediate sure-fire treatment. Whether educational efforts within the hospital and community can be effective is problematic. If, for example, Authoritarianism is characterologically imbedded, as the California group suggests (Adorno et al., 1950), no lecture series will dispel it. Perhaps, in hospitals at least, personnel selection on attitude factor scores may prove salutary for patients. But this should be preceded by empirical evidence concerning the relationships we all assume.

The present report describes the first step in a larger investigation. Work in progress includes assessment of opinion in the larger community, in the families of mentally ill veterans, and the relationship between personnel OMI factor scores and patient release rates over the 12 mental hospitals of the Veterans Administration's Psychiatric Evaluation Project.



## SUMMARY

A collection of 70 Likert-type opinion items, largely relevant to the mentally ill, was administered to most of the employees having frequent contacts with patients in two large, geographically widely separated Veterans Administration neuropsychiatric hospitals ( $N = 541$  and  $653$ ). The purpose of this investigation was to identify and measure the salient dimensions underlying these opinions and to begin the exploration of the construct validity of these dimensions by determining their relationships to the respondents' occupation, education, age, and sex. To this end, in each hospital separately, the item inter-correlations were subjected to centroid factor analysis followed by quartimax rotation. Scales were developed to measure each of the five factors identified in the analyses, and the resulting factor scores were related by analyses of variance to occupation, education, age, and sex. The results of all analyses in the two hospitals were the same in all essential regards; thus, the following conclusions apply to both:

1. Five salient opinion-attitude dimensions were identified. Therefore, attempts to work in this area with single scales (e.g., "pro-anti" mental patient, custodialism-humanism) oversimplify this domain. Further, correlations among factor scores for these dimensions are trivial or zero. The five factors were:

Factor A—Authoritarianism. This is clearly identified with the California F Scale and includes its authoritarian submission and anti-intracception combined with a view of the mentally ill as an inferior class requiring coercive handling. It accounts for about half the communal variance among the items, the other factors sharing the remaining half about equally.

Factor B—Benevolence. A kindly, paternalistic view towards patients whose origins lie in religion and humanism rather than science.

Factor C—Mental Hygiene Ideology. A positive orientation which embodies the tenets of modern mental health professionals and the mental hygiene movement whose leitmotif is "mental illness is an illness like any other."

Factor D—Social Restrictiveness. The central belief here is that the mentally ill are a threat to society, particularly the family, and

must therefore be restricted in their functioning both during and after hospitalization.

Factor E—Interpersonal Etiology. A circumscribed factor whose positive pole reflects the belief that mental illness arises from interpersonal experience especially deprivation of parental love during childhood.

2. Among the demographic variables, occupation and the closely related variable of education are substantially related to factor scores, particularly Factors A, B, and C. Education gives rise to curvilinear relationships with Factors B and C.

3. Age and sex show either zero or weak relationships with the factor scores.

## REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- BURT, C. *The factors of the mind*. New York: Macmillan, 1941.
- COHEN, J., & STRUENING, E. L. Factors underlying opinions about mental illness in the personnel of a large mental hospital. *Amer. Psychologist*, 1959, 14, 339. (Abstract)
- COHEN, J., & STRUENING, E. L. Attitudes toward the mentally ill of psychiatric hospital personnel as a function of occupation, education, sex, and age. *Amer. Psychologist*, 1960, 15, 417. (Abstract)
- CUMMING, ELAINE, & CUMMING, J. *Closed ranks*. Cambridge: Harvard Univer. Press, 1957.
- GILBERT, DORIS C., & LEVINSON, D. J. Ideology, personality, and institutional policy in the mental hospital. *J. abnorm. soc. Psychol.*, 1956, 53, 263-271.
- NEUHAUS, J. O., & WRIGLEY, C. The quartimax method: An analytical approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, 7, 81-91.
- NUNNALLY, J. C. The communication of mental health information: A comparison of experts and the public with mass media presentations. *Behav. Sci.*, 1957, 2, 222-230.
- STRUENING, E. L. The dimensions, distribution, and correlates of authoritarianism in a midwestern university faculty population. Unpublished doctoral dissertation, Purdue University, 1957.
- THOMSON, G. *The factorial analysis of human ability*. (5th ed.) New York: Houghton Mifflin, 1951.
- THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. Chicago Press, 1947.
- TRITES, D. K., & SELLS, S. B. A note on alternative methods for estimating factor scores. *J. appl. Psychol.*, 1955, 39, 455-456.
- TRYON, R. C. Reliability and behavior domain validity. *Psychol. Bull.*, 1957, 54, 229-249.
- TRYON, R. C. Cumulative communality cluster analysis. *Educ. psychol. Measmt.*, 1958, 18, 3-35.

(Received March 15, 1961)



## SOME EFFECTS OF PATERNAL ABSENCE ON MALE CHILDREN<sup>1</sup>

JOAN McCORD, WILLIAM McCORD, AND EMILY THURBER

*Sociology Department, Stanford University*

"That children are best reared in a home with two loving and understanding parents is so obvious as to need no statement" Dorothy Barclay (1959) has commented, typifying current opinion. This viewpoint is so prevalent that it comes close to heresy to question it. Although William Goode (1956), in his comprehensive study of divorce, points to the almost total lack of research on the effects of divorce on children, he concludes:

At every developmental phase of childhood, the child needs the father (who is usually the absent parent) as an object of love, security, or identification, or even as a figure against whom to rebel safely. . . . It would be surprising if the absence of the father had no effect on the child.

The same view prevails throughout social science. Few empirical studies of child development fail to include the words "intact homes" as a criterion of sample selection. It has long been the tradition to view anxiety as a primary outcome of father absence (Fenichel, 1945; Freud, 1953; Gardner, 1959). Such disorders as alcoholism, homosexuality, and totalitarian tendencies have been attributed to paternal absence (Meerloo, 1956). The high incidence of broken homes among the delinquent population has led to theories which might account for the apparent causative relationship (Burton & Whiting, 1960; Whiting, Kluckhohn, & Anthony, 1958).

In research comparing united homes with those in which the father is permanently or

temporarily absent, and in psychological and psychoanalytic theory concerning paternal absence, attention has been particularly centered on three areas of personality development: the extent to which the child develops a feminine as opposed to a masculine self-image, the intensity and type of anxiety which he experiences, and the probability of his engaging in antisocial behavior. In the following pages, we will examine various hypotheses in these areas as they relate to a (primarily) lower-class sample of boys. In the analyses, comparisons are made between boys raised in permanently broken homes and those in united homes. By varying the subgroups compared, the dynamic relationship between family disorder and abnormal behavior is assessed.

### METHOD

#### *Design of the Research*

During the 1930s, Richard Clark Cabot initiated the project, from which the subjects for this study of broken homes were taken, as an adjunct of an experimental program aimed at the prevention of delinquency in Cambridge and Somerville, Massachusetts (Powers & Witmer, 1951).

For an average period of 5 years, between the ages of 10 and 15, 255 boys<sup>2</sup> were observed at home, at school, and at play. Trained social workers, who

<sup>2</sup> Originally 325 boys had been included. Because of heavy case load, 65 boys were retired from the project in 1941, 5 additional boys were dropped because of their death or moving out of Massachusetts. The original sample was selected as follows: Teachers, police, and other officials recommended boys whom they believed showed signs of incipient delinquency. The Cambridge-Somerville Youth Study staff gathered information about them for the matching procedure (one boy to receive treatment and the other to be placed in a control group) so that the criteria of selection consisted in a willingness to participate and ability to find two boys with similar backgrounds in family structure, age, and "general personality." To avoid stigmatizing the boys in the project, an approximately equal number were added who were considered "normal" by the same authorities (again, equally divided between the treatment and the control groups).

<sup>1</sup> This research has been generously supported by the Ella Lyman Cabot Foundation, the Harvard Laboratory of Social Relations, and the National Institute of Mental Health (Grant M-2647).

Additional tables from this study have been deposited with the American Documentation Institute. Order Document No. 7059 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

visited the families approximately every other week, noted the behavior of the parents as well as the child. The counselors would appear unannounced, with a frequency which made it possible to observe the families at meals, during their leisure, in the midst of crises, and during their ordinary daily routines. They recorded their observations after each visit. Thus, running records were kept for 255 subjects between 1939 and 1945.<sup>3</sup>

In 1956 and 1957, trained researchers read each case record and rated the boy and his parents on a number of variables ranging from occupation and religion to affectional interaction. Interrater agreement, tested on a random sample, was high<sup>4</sup> and several factors point to the validity of the information obtained in this manner. Expected relationships which might have indicated a middle-class bias or operation of a halo effect were not found (e.g., the lower-class boys were not pictured as more aggressive and the brighter boy were not pictured as leaders). Most importantly, the categorized ratings of the case records yielded strong relationships to completely independent measures of social deviance among the subjects when they had become adults (McCord & McCord, 1960).

### Sample Characteristics

Among the 255 boys<sup>5</sup> in the study, 105 had lost one or both parents. Because we wished to focus on the effects of paternal absence, we dropped boys who were not living with their natural mothers (12 had lost both parents and 20 had lost their mothers) and the 18 who had step- or foster fathers. The remaining 55 boys from broken homes were living with their natural mothers; these were children whose fathers had died (24), deserted (8), been placed in mental hospitals (4), were serving long prison terms (3), or whose parents had been divorced or legally separated (16). The 150 boys whose natural parents were living together were used for the control group.

A number of studies have indicated that broken homes are associated with low socioeconomic status (Burgess, 1950; Hollingshead, 1950; Weeks, 1943). To the extent that social class affects personality development, this relationship between social class and family stability may lead to false conclusions regarding the effects of broken homes. Since the Cambridge-Somerville Youth Study centered upon the congested areas of these two cities, the entire sample

<sup>3</sup> Between 1955 and 1957, these subjects and a matched control group who had received no direct attention from the project were traced through the Massachusetts Board of Probation, mental hospitals, and various agencies dealing with alcoholism. It was found that the treatment program had no discernible effect upon criminality or alcoholism (McCord & McCord, 1959, 1960).

<sup>4</sup> The reliability of each of the ratings is fully discussed in *Origins of Alcoholism* (McCord & McCord, 1960).

<sup>5</sup> From 237 families.

TABLE 1  
PARENTAL DEVIANCE

Condition	Broken homes (N = 55)	United homes (N = 150)
Father deviant	30%	31%
Mother deviant	1	4
Both deviant	24	5
Neither deviant	45	60

had a strong lower-class representation. A comparison of fathers' occupations between broken and united homes within the sample showed slight (not statistically significant)<sup>6</sup> differences between the groups.

Various studies have indicated that Catholic families may be slightly more stable, although they seem to contribute more than their share of desertion cases (Bell, 1938; Monahan & Kephart, 1954). In our sample, records of the mother's religion showed that 65% of the boys raised in united homes and 64% of the boys in broken homes had been raised by Catholic mothers.

Theorists have also suggested that the wife may alter her behavior to compensate for her husband's absence. P. O. Tiller (1958) reports that Norwegian sailors' wives whose husbands were absent for extended periods of time exceeded matched mothers whose husbands were not absent in being overprotective and stressing obedience and politeness (in contrast to happiness and self-realization). In our sample, we did not find a significantly greater incidence of either maternal overprotection (31%/29%) or punitiveness (49%/44%) in the broken homes than in the united homes. Nor did we find significant differences between the mothers' attitudes toward their sons<sup>7</sup> in united homes and broken homes.

Two potentially important variables, however, strongly differentiated broken home boys from boys in united homes. William Goode (1956) reported that about a third of his sample of divorced women cited sexual or alcoholic deviance of their husbands as the primary cause of divorce. We found a significantly higher proportion of deviant (i.e., alcoholic, criminal, or promiscuous) fathers ( $p < .001$ ) and deviant mothers ( $p < .025$ ) among the boys from broken homes than among the boys whose parents were living together (see Table 1).

In addition, a significantly lower proportion of the boys in broken homes had immigrant fathers

<sup>6</sup> Throughout the research, the chi square test, two-tailed, was used when  $N > 30$  and the Fisher test, two-tailed, was used when  $N < 30$ . Differences were considered significant if  $p < .05$ .

<sup>7</sup> "Warm" mothers openly expressed their affection; "cold" mothers showed passive concern, but seldom demonstrated affection; "ambivalent" mothers displayed extreme variation between overt affection and overt rejection; and "rejecting" mothers cared little for their children or their welfare.



TABLE 2  
DISTRIBUTION OF BROKEN HOMES

Father	Mother	Age of boy at time of break		
		0-5	6-12	Over 12
Dead	Normal <sup>a</sup>	4	6	4
Living	Normal <sup>a</sup>	4	6	1
Dead	Abnormal <sup>b</sup>	3	3	4
Living	Abnormal <sup>b</sup>	7	11	2

<sup>a</sup> Warm, nondeviant.

<sup>b</sup> Deviant, cold, ambivalent, or rejecting.

( $p < .001$ ). Fifty-eight percent of the fathers in united homes, compared to 29% in broken homes, were immigrants.

To insure that the effects of these differences were not attributed to paternal absence, we matched each boy from a broken home to a boy similar in background whose parents were living together. Besides parental deviance and father's birthplace, the mother's attitude toward the boy, her disciplinary technique, her degree of control over her son, and the consistency of her discipline were used as criteria for matching.

We anticipated that paternal absence might have different effects under various conditions. Therefore, we divided the broken home boys on three dimensions:

1. The reason for the father's absence. The father's death might be presumed to have a different effect on the child than would his disappearance from the home after preliminary quarrels.
2. The age of the boy at the time when his father left. The child's age at the time of the break was divided roughly into preschool, preadolescent (or middle childhood), and adolescent.
3. The affectional relationship and stability (non-deviance) of the mother. Warm, nondeviant mothers were considered "normal." The distribution of the boys in broken homes on these three dimensions is shown in Table 2.

Since overt conflict probably precedes divorce and separation, and may have preceded desertion or death, the putative effects of broken homes may actually be the result of parental conflict. Ratings from direct observation of parental interaction were used to divide the boys whose parents were living together into two groups: the 30 whose parents quarreled constantly and were in overt conflict, and the 120 whose homes were relatively tranquil.

## RESULTS

### *Feminine Identification*

One of the most widely held beliefs about the effects of paternal absence is that male children will develop unusually strong feminine components in their personalities. Three

sets of ratings on the 205 boys in our study were used to test feminization in the father-absent group: homosexual tendencies, dependency, and lack of aggressiveness.<sup>8</sup> Although the trend of past evidence would suggest that father-absent boys would be relatively more feminine (Burton & Whiting, 1960; Leichthy, 1960; Lynn & Sawrey, 1959; Winch, 1949), more dependent (Stolz, 1954), and less aggressive (Bach, 1946; Sears, Pintler, & Sears, 1946), we found that neither homosexuality nor dependency differentiated significantly between the boys whose fathers were absent and those whose fathers were present and that the aggression scale was significantly related—but in the opposite direction from that predicted. (Eighty-seven percent of the broken home boys, as opposed to 67% of those from tranquil homes, were moderately or strongly aggressive.)

Since aggressive behavior may be considered as an exhibition of "masculinity," it seemed probable that those who were both aggressive and showed signs of feminine identification<sup>9</sup> were expressing an instability in sex role identification or defending against feminine identification. This combination of feminine-aggressive behavior (as compared to feminine-nonaggressive behavior) was found significantly more frequently among boys in broken homes than among boys in tranquil homes ( $p < .001$ ).

Since both parental conflict and paternal absence were related to feminine-aggressive behavior, it seemed likely that either parental

<sup>8</sup> Boys were considered to have strong homosexual tendencies if they played with dolls, sometimes wore dresses, frequently expressed the wish to be a girl, or were overtly homosexual. They were considered to be dependent if they showed an unusually strong desire for adult approval. Femininity and dependency were not significantly related to each other. A three-point scale of behavioral aggression, ranging from little to unrestrained, was used.

<sup>9</sup> Boys who evidenced high dependency or strong homosexual tendencies were classified as showing feminine identification. We hypothesized that feminine identification (with or without aggression) would arise from the "teasing" effect of an ambivalent nondeviant mother or from the combination of rejection from a stable father and affection from a stable mother; among the 22 boys whose parents were of these types, 77% evidenced feminine identification.



TABLE 3  
SEX ROLE BEHAVIOR

Sex role	Broken home (N=55)	Conflictful home (N=30)	Tranquil home (N=120)
Masculine <sup>a</sup>	49%	43%	58%
Feminine-nonaggressive	4	14	20
Feminine-aggressive	47	43	22

<sup>a</sup> Nine percent of the broken home boys, none of those in conflictful homes, and 13% of those in tranquil homes showed masculine role behavior but were not aggressive.

conflict or parental deviance (found in almost equal proportions among broken and conflictful homes)<sup>10</sup> might fully account for the difference. Neither of these explanations, however, fit the data. We reasoned that parental conflict would have been less among homes severed by death of the father; yet a higher proportion (58%) of the sons in these homes showed feminine-aggressive behavior (see Table 3).

To check whether the home milieu rather than paternal absence itself was responsible for the high rate of feminine-aggressive behavior, we used the group of boys with similar backgrounds in united homes.

Because the comparison with matched controls (see Table 4) showed higher feminine-aggressive behavior among broken home boys ( $p < .005$ ), the difference in sex role behavior could not be attributed simply to conditions which might have precipitated the family break.

Analysis of the father-absent boys provided a clue to their reasons for sex role conflict. We contrasted sons whose mothers were normal (affectionate and nondeviant) with those having mothers rated abnormal. In these two groups of boys, the age at which

<sup>10</sup> Immigrant families, too, were found in almost equal proportions among broken and conflictful homes, i.e., they were less likely to be either conflictful or severed.

TABLE 4  
SEX ROLE BEHAVIOR

Sex role	Father absent	Matched controls
Masculine	49%	45%
Feminine-nonaggressive	4	24
Feminine-aggressive	47	31

paternal absence began and the reason for such absence had different relationships to sex role behavior. These differences suggest that feminine-aggressive behavior has different origins in broken homes in which a normal, as opposed to an abnormal, mother had remained.

Whereas the child's age when his father left was of great importance among boys whose mothers were warm and nondeviant, it had slight relationship to feminine-aggressive behavior for boys raised by abnormal mothers (see Table 5).

Boys reared by normal mothers showed feminine-aggressive behavior only if their fathers left when the boys were between the ages of 6 and 12 ( $p < .01$ ). Since only 25% of their matched controls indicated sex role

TABLE 5  
PERCENTAGE WHO EXHIBITED FEMININE-AGGRESSIVE BEHAVIOR

	Boy's age when father left		
	0-5	6-12	Over 12
Normal mother	(N=8) 0	(N=12) 75 <sup>a</sup>	(N=5) 0***
Abnormal mother	(N=10) 70	(N=14) 50	(N=6) 50

<sup>a</sup> Twenty-five percent of matched group exhibited feminine-aggressive behavior ( $p < .05$ ).  
\*\*\*  $p < .01$ .

conflict of this type, the home milieu of these boys were apparently not responsible for their high rate. Studies of children's sex differentiated behavior give reason to believe that the years of middle childhood may be critical ones in the development of sex identification. In an early study, P. H. Furfey (1927) noted little sex differentiation in the play of 6-8 year olds, with increasing separation and differentiation after that age. Observations of recreational clubs at the Merrill-Palmer School in Detroit indicated that 5- and 6-year-olds seem to ignore sex as a basis for choosing play groups, but that sex segregation is almost complete for 10- and 11-year-olds (Campbell, 1939). Studies of friendship choices point to the same phenomenon (Moreno, 1934).

Previous research with father-separated samples whose mothers were probably "nor-

mal" by our criteria tend to point also to the importance of age at the time of separation. A study by Sears et al. (1946) found that early differences in sex role behavior between father-absent and father-present boys had begun to disappear by age 5. Bach (1946), however, reported evidence of feminization among 6-10 year olds whose fathers had been absent 1-3 years.

Early separation, as Sears et al. (1946) suggested, may result in sex typing delay—but both theirs and our evidence indicates that this effect is of relatively short duration: probably because the boy is able to find substitute role models during the period of sex identification. During the critical years of sex identification, perhaps because memory of the father interferes with adoption of a substitute model, loss of the father seems to have a more permanent affect on sex role identification. By age 12, the process of sex role identification is probably fairly complete, thus, explaining the absence of feminine-aggressive behavior among the older boys raised by normal mothers.

Among boys raised by abnormal mothers, age at the time of separation was of relatively minor importance in relation to feminine-aggressive behavior; death of the father (see Table 6), however, seemed to be highly productive of this type of confused sex role behavior ( $p < .05$ ).

One can argue that death of the father raises a conflict in the male child between his desire to replace the father and his denial of this desire; yet this theory does not explain the lower proportion among those whose mothers were affectionate ( $p < .05$ ) who showed feminine-aggressive behavior.

It seems reasonable to explain this type of sex role instability among boys exposed to cold or rejecting mothers in terms of dependency needs and their satisfactions: When resources for satisfaction of dependency needs are limited (as they would be in broken homes of this type), the child becomes both more dependent on this limited source and also more resentful of his dependency because it fails to bring satisfaction.<sup>11</sup> Thus, such children re-

<sup>11</sup> It seems likely that a relationship between maternal rejection and parental separation or divorce (see Newell, 1936) leads to disproportionate repre-

TABLE 6  
PERCENTAGE WHO EXHIBITED FEMININE-  
AGGRESSIVE BEHAVIOR

	Reason for father's absence	
	Death	Other
Normal mother	(N = 14) 36	(N = 11) 36
Abnormal mother	(N = 10) 90 <sup>a</sup>	(N = 20) 40*

<sup>a</sup> Ten percent of matched group exhibited feminine-aggressive behavior ( $p < .01$ ).

\*  $p < .05$ .

sponded to the conflict by being relatively dependent and feminine, and simultaneously behaving aggressively, in a compensatory masculine fashion.

### Anxiety

Although it has received less research attention, the belief that paternal absence results in anxiety is widespread. Specific research relating anxiety to paternal absence has yielded conflicting results. A number of studies have linked such various manifestations of anxiety as feelings of inferiority, poor school performance, immaturity, and tensions to paternal absence (Hardy, 1937; Lynn & Sawrey, 1959; Rouman, 1956; Stolz, 1954). Other studies have found no evidence of increased anxiety (Leichty, 1960; Rowntree, 1955; Russell, 1957). In an attempt to clarify this confusion in the literature, we tested three hypotheses derived from clinical theories.

*Hypothesis 1.* Father-separated boys should manifest many or intense fears because their heightened Oedipal desires cannot be brought to gratification (Freud, 1953), or because the child fears that his mother will desert him (Gardner, 1959). We found no confirmation of this hypothesis that loss of the father results in abnormal fears. Forty percent of the

sentation of this type of home in some studies of the effects of broken homes on sex role identification. Whiting (1961) suggested that cultures in which there are exclusive mother-child sleeping arrangements also tend to define the maternal role in terms which would be considered abnormal in our society (i.e., maternal rejection and promiscuity are common among them). It seems possible that the cross-cultural relationship found between father separation in infancy and evidence of sex role conflict is dependent upon the limited resources for satisfaction of dependency needs in these cultures.

TABLE 7  
PERCENTAGE WHO EXHIBITED SEX ANXIETY

Condition	Percentage
Broken home ( $N = 55$ )	47 <sup>a</sup>
Conflictful home ( $N = 30$ )	57
Tranquil home ( $N = 120$ )	27

<sup>a</sup> Forty-nine percent of the matched group exhibited sex anxiety.

broken home boys and 40% of those raised in tranquil homes gave evidence of abnormal fears (e.g., fear of the dark or excessive fear of bodily injury). Among boys reared in conflictful homes, 50% had abnormal fears. These negative results relating paternal absence to abnormal fears tend to confirm the findings of Rowntree (1955) for matched pairs of preschool children in Britain and Russell (1957) for matched pairs of school age children in America.

*Hypothesis 2.* Father-separated boys should have anxiety about sex; this should be particularly strong for those whose fathers have died (Fenichel, 1945). A number of boys expressed to their counselors their concern over achieving normal sexual relations or about their sexual adequacy; others publicly masturbated during periods of tension.<sup>12</sup> These boys were considered to be sexually anxious. Although a significantly higher proportion of those whose fathers were absent than of those whose homes were tranquil evidenced sex anxiety ( $p < .02$ ), roughly the same proportion of those whose parents were in open conflict were sexually anxious (see Table 7).

There was little variation within the father-absent group in the proportions who showed sex anxiety: 45% of those whose fathers were

<sup>12</sup> The two measures were significantly related in the sample ( $p < .01$ ).

TABLE 8  
ORAL TENDENCIES

	Broken home ( $N = 55$ )	Conflictful home ( $N = 30$ )	Tranquil home ( $N = 120$ )
Oral regression	22% <sup>a</sup>	13%	10%
Oral and anxious	15	23	16
Neither	63	64	74

<sup>a</sup> Thirteen percent of the matched group exhibited oral regression.

living, compared to 50% of those whose fathers had died; 54% of those whose mothers were rejecting and 41% of those whose mothers were affectionate evidenced sex anxiety. None of the four boys with affectionate deviant mothers evidenced sex anxiety. Although sex anxiety was prominent among boys raised without their fathers, the fact that 49% of the matched controls (compared to 47%) exhibited sex anxiety suggests that high sex anxiety may not be specifically related to paternal absence.

*Hypothesis 3.* Father-separated boys should show signs of regression (Fenichel, 1945). Thumb sucking, nail biting, excessive smoking, and constant playing with the mouth were used as behavioral signs of oral tendencies. Since these forms of behavior may also indicate general anxiety, only those who did not exhibit abnormal fears were classified

TABLE 9  
PERCENTAGE WHO SHOWED SIGNS  
OF ORAL REGRESSION

	Father absent	Matched controls
Normal mother	( $N = 25$ ) 8%	( $N = 25$ ) 20%
Abnormal mother	( $N = 30$ ) 33	( $N = 30$ ) 7**

\*\*  $p < .025$ .

as showing oral regression. Oral regression, though not oral anxiety, was found most frequently among the father-absent group. The relationship was not, however, strong enough to reject the possibility that it had occurred by chance (see Table 8).

As a further check, we examined oral regression in relation to normal and abnormal mothers among the father-absent boys. Although the proportion showing oral regression was not higher among the normal mother group than among those raised in tranquil homes (8%/10%), the comparison revealed a significantly higher proportion (see Table 9) showing signs of oral regression (33%) among those whose mothers were rejecting or deviant ( $p < .005$ ).

Rejection or deviance, with or without paternal absence, might have explained oral regression. Comparison with the matched group led to rejection of this hypothesis.



Reasoning that death of the father would most fully realize the Oedipal wish, we hypothesized greater regression among boys whose fathers had died. This hypothesis, too, was not supported. These comparisons indicate that paternal absence, probably following conflict, *in combination with* maternal deviance or rejection result in oral regression.

### Antisocial Behavior

The lay public as well as professional criminologists have linked broken homes to antisocial behavior. There seems to be general agreement that the proportion of broken homes among criminals is greater than that of the general population (Shulman, 1959). It was possible to use two measures of antisocial behavior for our sample. The counselors' reports of direct observations permitted ratings of primary reference groups during adolescence. Boys whose primary reference groups were delinquent gangs participated in behavior disapproved by the majority in their community. In 1955, court records for each of the subjects were obtained as an additional independent record of criminality; these traced the boys into adulthood. Those who had been convicted for a felony (or for a crime which would be a felony if the boy were an adult) were considered criminals.

There was little support for the theory that paternal absence led to delinquent gang activities. A significantly higher proportion of those boys whose parents continued to live together despite considerable overt conflict than *either* those whose parents were in little conflict ( $p < .01$ ) or those whose fathers were absent ( $p < .05$ ) were gang delinquents (see Table 10).

That parental conflict rather than paternal absence tends to result in gang delinquency is given further support by the fact that the

TABLE 11  
PERCENTAGE WHO BECAME CRIMINALS

Condition	Percentage
Broken home ( $N = 55$ )	36
Conflictful home ( $N = 30$ )	40
Tranquil home ( $N = 120$ )	22

older the boy at the time of the break, the more likely he was to become a gang delinquent. It should further be noted that a significantly higher proportion of those who had parent substitutes (34%) than of those who lived in tranquil homes had become gang delinquents. This latter group, it appears, is responsible for the apparently high rate of juvenile delinquency among the broken home population of the lower class—a correlation which has been erroneously attributed to the absence of a paternal model.<sup>13</sup>

Using convictions for felonies as a measure of antisocial behavior, the expected relatively high rate of criminality was found among the father-absent group (see Table 11). Tranquil homes produced a significantly lower proportion of criminals than did the father-absent homes and the conflictful homes ( $p < .025$ ).

Several findings point to the fact that the absence of a generally stable home environment, rather than the specific absence of the father, is related to criminality: (a) boys reared by parents who were in overt conflict were as likely to become criminals as boys from father-absent families; (b) the criminal rate among boys who had parent substitutes was identical (i.e., 36% became criminals) to that of the father-absent boys; (c) the criminal rate increased with an increase in the age of the boy at the time of the family break;

<sup>13</sup> With this theory in mind, we recomputed the Glueck (Glueck & Glueck, 1950) figures reported in *Unraveling Juvenile Delinquency*, breaking down the broken home boys into those who did and those who did not have parent substitutes. Recomputed, the Glueck figures no longer support the theory that broken homes as such are causally related to delinquency: Among their 500 delinquents, 72 were from broken homes without parent substitutes; among their 500 nondelinquents, 111 were from broken homes without parent substitutes. In contrast, 230 of the delinquents, compared to 60 of the nondelinquents, had substitute parents.

TABLE 10  
PERCENTAGE WHO HAD DELINQUENT  
REFERENCE GROUPS

Condition	Percentage
Broken home ( $N = 55$ )	20
Conflictful home ( $N = 30$ )	43
Tranquil home ( $N = 120$ )	18

and (d) none of the 13 father-absent boys cared for by warm nondeviant mothers whose fathers had not been deviant became criminals.<sup>14</sup>

### SUMMARY

Repeated direct observations of 205 boys and their families during a period of approximately 5 years of their early adolescence and court records for convictions for felonies were used to assess the effects of paternal absence upon boys. The sample, drawn from former members of the Cambridge-Somerville experiment, came from a lower-class, relatively deprived environment. The results of this study suggest the following conclusions:

1. Although feminine-nonaggressive behavior was negatively related to paternal absence, feminine-aggressive behavior appeared to be produced by paternal absence if the boy was between 6 and 12 when his father left, or the mother was deviant or rejecting (especially if the father had died).
2. No support was found for the theory that paternal absence leads to abnormal fears.
3. Intense sexual anxiety was found among almost half of the boys who had lost their fathers. Yet this anxiety seemed to be a response to a generally unstable environment rather than to paternal absence per se.
4. Oral regression was related to father-absence only among those whose mothers were deviant or rejecting.
5. Gang delinquency was found to be unrelated to paternal absence, although it did occur more frequently in broken homes in which the father or mother had been replaced by substitutes. In fact, the proportion of gang delinquents among boys whose parents quarreled but remained together was significantly higher than among those whose fathers were absent.
6. The relationship between criminality and paternal absence appears to be largely a result of the general instability of broken homes rather than of paternal absence in itself.

The evidence drawn from this sample indicates that many of the effects often presumed

to result from paternal absence can, largely, be attributed to certain parental characteristics—intense conflict, rejection, and deviance—which occur more commonly in broken families.

### REFERENCES

- BACH, G. R. Father-fantasies and father typing in father-separated children. *Child Developm.*, 1946, 17, 63-79.
- BARCLAY, DOROTHY. When one parent plays the double role. *N. Y. Times Mag.*, 69, April 5, 1959.
- BELL, H. *Youth tell their story*. Washington, D. C.: American Council on Education, 1938.
- BURGESS, E. W. Predictive methods and family stability. *Ann. Amer. Acad. Pol. Soc. Sci.*, 1950, 272, 47-52.
- BURTON, R. V., & WHITING, J. W. M. The absent father: Effects on the developing child. (Rev.) Paper read at American Psychological Association Convention, Chicago, 1960.
- CAMPBELL, E. H. The social-sex development of children. *Genet. psychol. Monogr.*, 1939, 21, 461-552.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- FREUD, S. Three essays on sexuality. In, *Standard edition*. Vol. VII. (Originally published 1905) London: Hogarth, 1953.
- FURFEY, P. H. Some factors influencing the selection of boys' chums. *J. appl. Psychol.*, 1927, 11, 47-51.
- GARDNER, G. E. Separation of the parents and the emotional life of the child. In S. Glueck (Ed.), *The problems of delinquency*. Boston: Houghton Mifflin, 1959. Pp. 138-143.
- GLUECK, S., & GLUECK, ELEANOR T. *Unraveling juvenile delinquency*. Cambridge, Mass.: Harvard Univer. Press, 1950.
- GOODE, W. J. *After divorce*. Glencoe, Ill.: Free Press, 1956.
- HARDY, M. C. Aspects of home environment in relation to behavior at the elementary school age. *J. juv. Res.*, 1937, 21, 206-225.
- HOLLINGSHEAD, A. B. Class differences in family stability. *Ann. Amer. Acad. Pol. Soc. Sci.*, 1950, 272, 39-46.
- LEICHTY, MARY. The absence of the father during early childhood and its effect upon the Oedipal situation as reflected in young adults. *Merrill-Palmer Quart.*, 1960, 6, 212-217.
- LYNN, D. B., & SAWREY, W. L. The effects of father-absence on Norwegian boys and girls. *J. abnorm. soc. Psychol.*, 1959, 59, 258-262.
- MCCORD, W., & MCCORD, JOAN. *Origins of crime*. New York: Columbia Univer. Press, 1959.
- MCCORD, W., & MCCORD, JOAN. *Origins of alcoholism*. Stanford, Calif.: Stanford Univer. Press, 1960.
- MEERLOO, J. A. M. The father cuts the cord: The role of the father as initial transference figure. *Amer. J. Psychother.*, 1956, 10, 471-480.

<sup>14</sup> Nine of the 10 father-absent boys whose mothers were both rejecting and deviant had been convicted for felonies.

- MONAHAN, T. P., & KEPHART, W. M. Divorce and desertion by religious and mixed-religious groups. *Amer. J. Sociol.*, 1954, 59, 454-465.
- MORENO, J. L. *Who shall survive?* Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934.
- NEWELL, H. W. The psycho-dynamics of maternal rejection. *Amer. J. Orthopsychiat.*, 1936, 6, 576-588.
- POWERS, E., & WITMER, HELEN. *An experiment in the prevention of delinquency.* New York: Columbia Univer. Press, 1951.
- ROUMAN, J. School children's problems as related to parental factors. *J. educ. Res.*, 1956, 50, 105-112.
- ROWNTREE, GRISELDA. Early childhood in broken families. *Popul. Stud.*, 1955, 8, 247-263.
- RUSSELL, I. L. Behavior problems of children from broken and intact homes. *J. educ. Sociol.*, 1957, 31, 124-129.
- SEARS, R. R., PINTLER, M. H., & SEARS, PAULINE S. Effects of father-separation on preschool children's doll play aggression. *Child Developm.*, 1946, 17, 219-243.
- SHULMAN, H. M. The family and juvenile delinquency. In S. Glueck (Ed.), *The problems of delinquency.* Boston: Houghton Mifflin, 1959. Pp. 128-136.
- STOLZ, LOIS M., ET AL. *Father relations of war-born children.* Stanford, Calif.: Stanford Univer. Press, 1954.
- TILLER, P. O. Father-absence and personality development of children in sailor families: A preliminary research report. *Nord. Psykol.*, 1958, Monogr No. 9.
- WEEKS, A. H. Differential divorce rates by occupations. *Soc. Forces*, 1943, 22, 334-337.
- WHITING, J. W. M. Paper read to Graduate Colloquium in Psychology, Stanford University, January 19, 1961.
- WHITING, J. W. M., KLUCKHON, R., & ANTHONY, A. The function of male initiation ceremonies. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (3rd ed.) New York: Holt, 1958. Pp. 359-370.
- WINCH, R. F. The relation between the loss of a parent and progress in courtship. *J. soc. Psychol.*, 1949, 29, 51-56.

(Received March 27, 1961)



## CLINICAL PATTERNS OF DEFENSE AS REVEALED IN VISUAL RECOGNITION THRESHOLDS<sup>1</sup>

DONALD T. SHANNON

*University of Illinois*

Recently, interest in the perceptual defense hypothesis has been subsiding. It would appear that the storm of questions and criticisms raised in numerous experimental and theoretical analyses has discouraged further study of the concept. Questions of the effects of variables such as familiarity, set, recency, and frequency of occurrence (Howes & Solomon, 1950; Postman, 1953; Postman, Bronson, & Gropper, 1953; Solomon & Postman, 1952), as well as what Eriksen (1958) calls the "bugaboo of unconscious perception," have all but obscured the experiments which have demonstrated that psychological defense mechanisms are detectable in perceptual recognition behavior (Chodorkoff, 1954; Eriksen, 1951, 1952a, 1952b, 1954; Eriksen & Browne, 1956; Eriksen & Davids, 1955; Lazarus, Eriksen, & Fonda, 1951). These studies cannot be dismissed on the grounds of poor methodology or design. Although they have not directly contributed to our understanding of the perceptual processes involved, they do represent an important body of information which is useful to our understanding of defensive behaviors. Even before we know how the individual defends himself against threatening stimuli, it can be argued that his defensive responses are in themselves important data, particularly in the clinical study and description of the individual.

Eriksen's (Eriksen & Browne, 1956) argument that the individual's perceptual processes and his verbal recognition responses can

be distinguished and studied separately lends an added rationale for the continued study of defensive behaviors by perceptual recognition experiments.

The present study attempts to show that certain groups of hospitalized psychiatric patients, who utilize different defensive patterns in dealing with their conflicts and anxieties, respond differentially in their reported recognition of conflict relevant stimuli.

Consistent individual differences in response to situations where anxiety and self-esteem threat is experienced have been reported in numerous previous studies. Several such studies (Carlson, 1954; Eriksen & Browne, 1956; Eriksen & Davids, 1955) have demonstrated that such differences can be related to scores on the *Hy* and *Pt* scales of the MMPI, two scales interpreted as measures of ways subjects handle their anxieties. Eriksen and Davids (1955) report that subjects with high *Pt* scores and low *Hy* scores are clinically rated as using less repression than subjects with low *Pt* and high *Hy* scores. They suggest that the research relating MMPI scales to differences in response to anxiety distinguishes between persons on the same behavior variables that are clinically found related to defensive reactions. The present study extends the demonstrated differences between such diagnostic groups (*Hy* and *Pt*) to a broader diagnostic classification.

Many of the previous studies of perceptual defensive behaviors were conducted with normal subjects (Blum, 1954; Chodorkoff, 1954; Neel, 1954). It is believed that the use of hospitalized psychiatric patients enables us to study the effects of more extreme degrees of defensive behaviors on recognition thresholds. Further, it is important to make sure that the stimuli presented are indeed anxiety arousing for the individual subjects.

The two concepts involved in this study of perceptual recognition behaviors are "percep-

<sup>1</sup> This article is adapted from a dissertation submitted to the Department of Psychology and the Committee on Graduate Study of Stanford University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to express his appreciation to the members of the doctoral committee: Kenneth B. Little, Chairman; Quinn McNemar and Sanford Dean, for their help in the planning of the study; and to the staff and the patient group of the Veterans Administration Hospital, Palo Alto, for their participation in the study.

tual sensitization" and "perceptual defense." As first formulated by Postman, Bruner, and McGinnies (1948), *perceptual sensitization* is the lowering of the reported recognition threshold for an object or concept which is acceptable to the organism, or which can be used to avoid or reduce discomfort and maintain well-being. *Perceptual defense*, on the other hand, is the elevation of recognition thresholds for stimuli which would result in discomfort, or which would threaten the individual's sense of well-being. Both phenomena are assumed to be adjustive and purposive in the general life functioning of the individual.

On the basis of these definitions, one would expect that all individuals would rely on perceptual defense for all actual or potential anxiety producing ideas of objects. It is a common clinical observation, however, that certain individuals tend to respond to threatening stimuli in an immediate and direct fashion. Rather than reacting by delayed recognition or avoidant behavior, they seem to be on the lookout for them. In such instances, presumably there is a lowering of thresholds for anxiety producing stimuli, and thus sensitization toward them, rather than perceptual defense. This seems to be particularly true when the early recognition of the stimuli permits the individual to avoid the associated pain or discomfort. Perceptual defense would seem to occur more often when anticipatory recognition of stimuli will not reduce or help avoid discomfort. On the basis of such reasoning, Rosen (1954) was able to produce alternation in perceptual recognition thresholds as a function of the utility of the increased or delayed perceptual response in avoiding painful stimuli. The results of his study generally substantiated the hypothesis of the utilitarian function of perceptual recognition behavior in maintaining the individual's well-being and equilibrium.

The logic of the reasoning concerning the adjustive utility of perceptual recognition behavior led directly to the present attempt to relate perceptual sensitization and defense to defensive reactions observed in clinical practice. On the basis of both theory and observational experience, it was possible to develop predictions of differences in recognition

thresholds of persons utilizing different defensive patterns.

Three generalized defensive patterns were identified: Externalization, Internalization, and Acting-out. This classification of defensive patterns most nearly resembles Rosenzweig's (1945) grouping of reactions to aggressive, and hostile impulses. In the present study, individuals who use the externalization defensive patterns (Externalizers) tend to reduce anxiety by denying the personal origin of the conflict or impulse, and find objects or concepts in their environment to attribute or project the motivation onto. Individuals using the internalization defensive pattern (Internalizers) tend to reduce their anxiety by denial of the personal relevance of the conflict, but avoid even the projected or displaced expression of it. Individuals utilizing the third defensive pattern of acting-out, reduce anxiety associated with impulses or conflicts by immediately discharging them, usually symbolically, in verbal and nonverbal behavior. This group is similar to the Externalizers in the direction of expression, but they do not necessarily involve active denial of the personal relevance or origin of the impulse or conflict.

The following hypotheses were formulated:

*Hypothesis 1.* Individuals utilizing the externalization defensive pattern will exhibit perceptual sensitization to anxiety producing stimuli. The specific prediction is that Externalizers will report a lower recognition threshold for conflict relevant stimuli than for matched neutral stimuli.

*Hypothesis 2.* Individuals using the internalization defensive pattern will exhibit perceptual defense against threatening stimuli. The prediction is that Internalizers will report a higher recognition threshold for conflict relevant stimuli than for matched neutral stimuli.

*Hypothesis 3.* Individuals who typically act-out their conflicts and impulses directly, will exhibit perceptual sensitization for threatening stimuli. The prediction is that Acters-outers will report lower recognition thresholds for conflict stimuli than for matched neutral stimuli. The expectation here, however, is that Acters-outers will show less differential in their



recognition thresholds between conflict and neutral stimuli, than will the Externalizers.

A fourth corollary hypothesis tested is that the difference in the recognition thresholds would be greater between stimuli related to the individual's primary problem areas and matched neutral stimuli than between those stimuli related to secondary problem areas and their matched neutral stimuli.

### METHOD

Subjects were 51 adult male hospitalized psychiatric patients between the ages of 21 and 45 years (median age of 32 years), selected by the following criteria: normal vision, demonstrated ability to read and write English, normal intelligence, willingness to cooperate with experimental procedures, and the availability of a complete clinical history, including psychological tests, case history data, and clinical observations.<sup>2</sup>

Selection and classification of each subject into one of the three categories was based on consensus judgment of the ward psychiatrist, ward nursing staff, group or individual therapist for subjects undergoing treatment, and at least one psychiatric aide. These judges were contacted separately and asked to submit names of patients who were known to have their major conflicts in one of the following three problem areas: control of aggressive, hostile impulses; expression of sexual impulses; and feelings of dependency, and who typically defended themselves against one or more of these conflicts by using one of the three defensive patterns studied. Following this, the judges were seen in a group to discuss and verify their separate clinical evaluations of patients who met the selection criteria. Using this procedure, three groups of 17 patients each, were obtained for the final sample.<sup>3</sup>

<sup>2</sup> Comparison of the three groups with respect to psychiatric diagnoses disclosed the following distribution: Externalizers ( $N=17$ ): Schizophrenic Reaction (Paranoid), 8; Schizophrenic Reaction (Undifferentiated), 6; Inadequate Personality, 1; Phobic Reaction, 1; Anxiety Reaction, 1; Internalizers ( $N=17$ ): Schizophrenic Reaction (Paranoid), 11; Schizophrenic Reaction (Catatonic), 1; Schizophrenic Reaction (Mixed), 1; Schizophrenic Reaction (Undifferentiated), 4; Acter-outers ( $N=17$ ): Schizophrenic Reaction (Paranoid), 5; Schizophrenic Reaction (Hebephrenic), 2; Schizophrenic Reaction (Catatonic), 1; Schizophrenic Reaction (Simple), 1; Schizophrenic Reaction (Undifferentiated), 6; Inadequate Personality, 1; Anxiety Reaction, 1.

<sup>3</sup> During the selection and classification procedures, it became apparent that the subjects were seen by the judges as experiencing some degree of conflict in all three areas selected. Further investigating resulted in the impression that the three conflict areas selected essentially would apply to all the patient population. It was decided to treat the conflict

While there is no quantitative estimate of the reliability of this evaluation and selection procedure, the inclusion of only those patients about whom there was complete agreement by the judges insured a meaningful clinical classification of subjects.

### Measures

Two types of visual stimuli were used in the recognition measures. The first type consisted of 30 words, 15 conflict relevant and 15 neutral. The 15 conflict words consisted of three sets of five words, each set relating to one of the three conflict areas. These conflict words were selected on the basis of the pooled ratings of 10 clinical psychologists. The 15 neutral words were selected from the Thorndike-Lorge (1944) tables as having the same frequency of usage as the conflict words. In addition, only those neutral words which contained the same number of letters and resembled the conflict words in configuration were selected. Conflict words which were not listed in the tables were assigned the lowest frequency category.

Each stimulus word was printed in ink on clear plastic and made into  $3\frac{1}{2} \times 4$ " slides for presentation. In the word stimuli test, each conflict relevant word was preceded by a neutral word. Except for this limitation, the order of presentation of words was randomly determined and different for each subject within a given group. However, the same order of presentation was used in all three groups. The matched conflict and neutral words are set forth in Table 1.

The second set of visual stimuli consisted of 16 line drawings. Eight drawings depicted scenes which were judged neutral by the same group of clinical psychologists who had selected the conflict words and the remaining eight drawings consisted of two sets of four pictures each, judged to represent the areas of sex and aggression. Due to the lack of rater agreement on pictorial depiction of dependency, drawings illustrating this conflict area were not included in the test. The neutral and conflict pictures were matched in so far as all scenes involved two people, with males and females appearing the same number of times in the conflict and neutral pictures. These drawings were photographed, and the positive negatives made into  $3\frac{1}{2} \times 4$ " slides. In the picture stimuli test, each conflict relevant picture was preceded by a neutral picture.

In order to determine the effect of familiarity on picture recognition scores, and to a certain extent the effect of set, all subjects were shown one-half (eight) of the pictures prior to the experimental testing. The particular set of eight pictures was varied so that no consecutive subjects within one of the three groups were exposed to the same set. This procedure of familiarizing one-half of the pictures con-

stimuli as representing conflict areas of varying intensity for all subjects. The essential function of the judge's evaluation was to designate a particular conflict area as primary or secondary for each subject.



TABLE 1  
LIST OF STIMULUS WORDS

Conflict words	Neutral words
STRANGLE	STRAGGLE
STAB	STEW
SMASH	SHORT
SHOOT	SHEEP
KILL	KING
MOTHERLY	MOLECULE
BEGGING	BREATHS
CLINGING	CLUSTERS
HELPLESS	HIGHWAYS
NURSING	NESTING
COCK <sup>a</sup>	COOT
CUNT <sup>a</sup>	CURD
WHORE	WHELP
ERECTION	EYESIGHT
PENIS	PENCE

Note.—The conflict relevant and neutral words were matched for frequency as determined from Thorndike-Lorge (1944) word count.

<sup>a</sup> Not listed in Thorndike-Lorge count; assigned lowest frequency count.

sisted of showing the subjects the 15" × 22" original drawings of the scenes and giving the criterion description. The subject was told that he would be shown the same pictures later on slides, along with others which he had not seen.

Stimulus slides for both words and pictures were presented from a projection lantern. Attached was a rheostat to increase and systematically control the illumination of the slides. The slides were shown on a 40" × 46" cloth screen at a distance of 8' from the subject. At this distance, the letters in the words were 3" high and 2.5" wide; the picture image was 28" high and 40" wide.

#### Procedure

The actual testing session lasted approximately 1.5 hours. After the subject was acquainted with the set-

ting, he was presented with three practice word slides and three picture slides just prior to the presentation of the experimental word or picture tests. This procedure was based upon the findings in a pilot study that there was a noticeable drop in the recognition threshold between the first and second word or picture stimulus, but the recognition threshold tended to be stabilized beginning with the third stimulus.

In the experimental tests, each stimulus was presented once, using the procedure of increasing the illumination of the slide from a previously established level of illumination well below the recognition threshold of any of the pilot subjects, to the level where correct identification was reported. The subject was instructed to report his identification of the stimulus as soon as he could read or see it, "even before he was sure," and "even if the word or picture seemed foolish or unusual, because some were unusual." The recognition threshold score for the stimuli was taken as the number of volts registered on the rheostat when the subject correctly identified the stimulus.

In certain of the analyses, differences scores (DS) were used. These scores were the deviation of the reported recognition threshold for conflict stimuli from neutral stimuli. The use of this intraindividual statistic permitted the control of individual differences in general perceptual acuity.

#### RESULTS

Table 2 presents the difference between the mean recognition scores for conflict and neutral words and pictures for the Externalizers, Internalizers, and Acter-outers. Also presented are the standard deviation for the set of recognition scores, and the standard error of the differences between the correlated mean scores. A minus DS indicates that reported recognition of conflict stimuli is at a lower threshold than reported recognition for neu-

TABLE 2  
DIFFERENCE BETWEEN MEAN RECOGNITION SCORES FOR CONFLICT (C) AND NEUTRAL (N) STIMULI

Group	Conflict		Neutral		Differences		t
	M	$\sigma$	M	$\sigma$	M	$\sigma$ MD	
Words							
Externalizers	58.83	3.43	67.85	7.46	-9.02	1.87	4.82***
Internalizers	69.51	13.27	61.73	8.18	7.78	3.75	2.08*
Acter-outers	57.80	6.51	64.60	7.11	-6.80	.81	8.39***
Pictures							
Externalizers	71.91	7.02	88.39	13.27	-16.48	-2.63	6.27***
Internalizers	86.74	10.57	75.10	6.20	11.64	1.96	5.93***
Acter-outers	74.01	14.76	82.88	11.62	-8.87	-2.66	3.33**

\* Significant at .02 level.

\*\* Significant at .01 level.

\*\*\* Significant at or beyond .001 level.

TABLE 3  
MEAN DIFFERENCE SCORES FOR CONFLICT RELEVANT WORDS AND PICTURES

Group	Words				Pictures		
	Sex	Aggressive	Dependent	M	Sex	Aggressive	M
Externalizers	-8.48	-7.94	-7.25	-7.89	-19.29	-12.09	-15.69
Internalizers	10.46	7.72	7.22	8.46	10.31	15.48	12.89
Acter-outers	-5.73	-6.53	-7.00	-6.42	-12.45	-5.99	-9.22
M	-1.25	-2.25	-2.34		-7.14	-.87	

Note.—These scores were computed as the deviation of the recognition threshold for conflict stimuli from neutral stimuli.

tral stimuli, while a positive DS indicates that the conflict stimuli have a higher reported recognition threshold.

All three major hypotheses are confirmed. Externalizers and Acter-outers report lower recognition thresholds, and Internalizers report higher recognition thresholds for both the conflict words and pictures.

Recognition threshold scores for the conflict words and pictures correlate .69 for the three groups combined. Recognition scores for the neutral words and pictures correlate .47. Both coefficients are significant beyond the .001 level, and support the conclusion that the two types of visual stimuli evoke similar recognition behavior in the individuals studied.

Certain additional questions concerning the recognition difference scores associated with these three patterns of defensive behavior require examination. The first is whether the three groups differ among themselves in their differential response to conflict and neutral stimuli. The second question, of interest in the clinical field, is whether the three groups vary significantly in their response to the conflict and neutral words and pictures, and the third is whether the specific types of defensive reactions are significantly associated with particular conflict stimuli.

To consider these questions, a pseudo-triple-classification analysis of variance (McNemar's, 1955, Case XVII) of the difference scores of the three groups for the words, and for the picture stimuli was made. The results of the analysis with the word stimuli indicate that only the variation among the groups is significant. The differences in the mean difference scores for the three types of conflict words (sex, aggression, dependency) are not

significant. The interaction between types of defensive reactions and types of conflict words is similarly nonsignificant. Table 3 presents the mean difference scores for the three groups on the words and pictures related to the conflict areas.

From the results presented in Table 2 we would expect the difference between the DSs of the Internalizers and Externalizers, and between the Internalizers and Acter-outers, to be significant. It was of interest, however, to test the hypothesized difference in the size of the mean DSs for Externalizers and Acter-outers, since the same direction was predicted for both groups. The difference between the DSs of these two groups for the word stimuli is nonsignificant ( $t = .67$ ); the difference between their DSs for the picture stimuli, however, is significant at the .02 level ( $t = 2.22$ ).

The results of the triple analysis of variance of the picture DSs show a similar significant difference among the mean DSs of the three groups. Unlike the results with the word stimuli, the difference in the mean DSs for the two types of conflict pictures (sex and aggression) is significant for the combined groups. However, the interaction between types of defensive reactions and types of conflict pictures is nonsignificant. Moreover, Externalizers and Acter-outers differ significantly in the magnitude of their DSs for conflict relevant pictures.

Analysis of the effects of the familiarization procedure on the reported recognition thresholds of conflict and neutral pictures indicates that familiarizing subjects to one-half of the pictures in the pretesting procedures has the effect of significantly lowering the reported recognition scores in the experimental test

sessions for both the familiarized conflict relevant and neutral pictures, for all groups. However, the mean difference between conflict and neutral pictures remains significant beyond the .001 level for the three groups under both conditions. One may conclude, then, that while the familiarization procedure tends to lower the recognition scores for both conflict and neutral pictures, the direction and the magnitude of the difference in recognition thresholds for conflict and neutral pictures for the three groups is essentially unaffected.

In addition, an analysis was made of the possible effects involved in the subject becoming acquainted with the type of task required by the experiment. The question arises whether exposing the subject to one or the other tasks first (word or pictures), has the effect of lowering the recognition threshold for the other type of stimuli, presented subsequently. This question not only has to do with familiarization effects, but also involves the possible influence of set, although the exact process cannot be specified.

The mean recognition scores for conflict and neutral words and pictures under the two experimental procedures of presenting words first or presenting pictures first were analyzed by four double classification analyses of variance. The first analysis was of the conflict word recognition scores under two orders of presentation. The results indicate that presenting the pictures first does not significantly effect the recognition scores for the conflict words for the three groups. The second analysis was of neutral word recognition scores. Results of this analysis indicate that presenting the pictures first has the effect of significantly raising the recognition scores for the neutral words for the Externalizers and Internalizers, but does not significantly alter the recognition scores of neutral words for the Acter-outers. This interaction effect was significant at the .01 level. The third analysis was of the recognition scores for conflict pictures. Presenting the words first significantly raises the reported recognition of conflict pictures for both the Externalizers and Acter-outers, but lowers the reported recognition of conflict pictures for the Internalizers. This interaction effect is significant at the .001 level.

The fourth double classification analysis of variance was of the neutral pictures recognition scores. In this case presenting the words first has the effect of significantly raising the recognition scores for neutral pictures for the Externalizers and Acter-outers, but the reverse effect of significantly lowering the recognition scores of neutral pictures for the Internalizers. The interaction effect here is significant beyond the .001 level.

In addition to the main predictions, the following secondary prediction was tested: subjects will report larger difference scores between recognition of their primary conflict area and neutral stimuli, than between their secondary conflict area and neutral area. Because only two conflict areas were used in the picture test (sex, aggression) only subjects who had been classified as having one of these as their primary conflict were used in the analysis of the mean difference scores for picture stimuli.

No significant differences are found between recognition scores of words related to primary and secondary conflict areas. With picture stimuli, only one significant difference is found and that contrary to prediction. Externalizers show a larger difference between secondary conflict area and neutral pictures than for the primary-neutral comparison.

#### DISCUSSION

The results obtained support the hypothesis that groups of individuals using different types of clinical defensive patterns report different perceptual recognition thresholds for conflict related stimuli. The differences obtained and the predicted direction of the differences were not altered when the effects of familiarity with the stimuli and with the task were considered. Previous studies (Chodoroff, 1954; Postman, 1953) have described the occurrence of perceptual defense and/or perceptual sensitization in individuals but questioned whether such perceptual patterns were true for groups as a whole. The results reported here, while in no way contradictory of the interpretation of these phenomena as "individual" suggest, however, that they may be "general" when persons are categorized on the basis of their clinical defensive reactions. It was primarily the empirical findings of in-



dividual "exceptions" in these earlier studies which led to the present attempt to relate types of perceptual reactions to clinical types of defensive reactions.

The finding that the three groups did not vary significantly in their differential response to the three types of conflict words suggests a generality of defensive reaction to conflict stimuli. This is in contrast with the results reported by Kurland (1954) who found a lack of generality in defensive reactions of hospitalized patients to conflict words. However the present results are consistent with the general clinical impression that hospitalized psychiatric patients respond in most anxiety producing situations with a more total, "massive," generalized defensive reaction. The failure to find a significant association of types of conflict words with types of defensive reactions (considering the question concerning the relationship between conflict areas and different defensive reactions) was not unexpected since the particular defensive reactions studied here are ones which cut-across conflict areas.

The differences in the combined group's response to pictures portraying sexual and aggressive activities suggests that depiction of sexual behavior elicits a stronger defensive reaction generally. Aggression is more commonly depicted and expressed in our motion pictures and television media than blunt portrayal of sexual activity. This difference in social prohibitions, however, seems less prominent in written descriptions of these behaviors, which may help explain the failure to find differences in recognition of sex and aggression word stimuli.

The finding that Externalizers and Acters differ significantly in the magnitude of their difference scores for conflict pictures (Table 3) suggests that externalization is associated with greater sensitivity to conflict stimuli, and in a sense, less distortion, than the acting-out defensive reaction. In distinguishing between these two defensive patterns however, the acting-out pattern seems to involve less distortion in that conscious denial of impulses or anxiety is less characteristic, whereas active denial of the impulse is an important characteristic of externalization. The major characteristic of the acting-out pattern

of defense is lack of inhibition of the expression of the impulse, not the distortion of it.

The results of the analyses of the effects of the order of presentation of tasks seem inexplicable. No hypothesis examined was found satisfactory. The only clear implication of order effects is that no blanket statement can be made concerning the effects of familiarity on recognition thresholds without considering the defensive reaction pattern of the individual. Postman (1953) has made the statement that equal familiarity with stimuli tends to diminish differential recognition thresholds. The results reported here, however, show that familiarity tends to raise and/or lower recognition thresholds, depending on the defensive reaction characteristic of the individuals. Perhaps with normal subjects familiarity has the general effect of lowering recognition thresholds, but it appears that such an effect is not present in hospitalized psychiatric groups.

Five out of the six separate tests fail to support the secondary hypothesis that individuals respond differentially to stimuli related to their primary conflict areas and to their secondary conflict areas. There seems to be no explanation for the finding that one of the six tests was significant at the .05 level (but in the opposite direction predicted) other than chance. The failure to support the predicted differential response to primary and secondary conflict area stimuli is consistent with previously discussed finding that hospitalized patients show a generality in defensive reactions to all threatening stimuli.

Classification of psychiatric patients in terms of their patterns of recognition of threatening stimuli and/or interpersonal stress would seem clinically useful and important. The defensive patterns used here cut-across the lines of conventional diagnostic groups. They are descriptive of the direction of expression of defensive responses in individuals, and may be useful in studies of differential patient response to various types of reinforcement and psychiatric treatment. Perhaps most important, such a classification scheme is based on clinically observable behaviors. Ullmann (1960) has recently demonstrated the feasibility of such a classification, based on case history material, and has developed a rating scale for such purposes. In addition,

he has derived and cross-validated an MMPI scale which differentiates such groups.

### SUMMARY

This investigation was concerned with the visual recognition behaviors of individuals using different patterns of clinical defenses. It was hypothesized that the defensive patterns of *externalization* and *acting-out* would be associated with perceptual sensitization for threatening stimuli, whereas the defense pattern of *internalization* would be associated with perceptual defense against threatening stimuli. It was also hypothesized that patients would show differential recognition behaviors to primary and secondary conflict area stimuli.

Words relating to sex, aggression, and dependency, and pictures portraying sexual and aggressive behaviors were the two types of visual stimuli used for the experimental measures. Reported visual recognition was measured by the level of illumination required for identification.

It was predicted that clinically defined Externalizers and Acter-outers will report lower recognition thresholds, and Internalizers will report higher recognition thresholds for conflict stimuli than for matched neutral stimuli. The results confirm these predictions. The prediction that patients will show differential recognition responses to primary and secondary conflict area stimuli was not confirmed.

### REFERENCES

- BLUM, G. S. Perceptual defense revisited. *Amer. Psychologist*, 1954, 9, 335-336.
- CARLSON, V. R. Individual differences in recall of word-association test words. *J. Pers.*, 1954, 23, 77-87.
- CHODORKOFF, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, 49, 508-512.
- ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, 46, 557-564.
- ERIKSEN, C. W. Defense against ego threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, 47, 230-235. (a)
- ERIKSEN, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952, 44, 442-446. (b)
- ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, 61, 175-182.
- ERIKSEN, C. W., & BROWNE, C. T. An experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, 52, 224-230.
- ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, 50, 135-137.
- HOWES, D. H., & SOLOMON, R. L. A note on McGinnies "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, 57, 229-235.
- KURLAND, S. H. Lack of generality in defense mechanism as indicated in auditory perception. *J. abnorm. soc. Psychol.*, 1954, 49, 173-177.
- LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, 19, 471-482.
- MCNEMAR, Q. *Psychological statistics*. (2nd ed.) New York: Wiley, 1955.
- NEEL, ANN F. Conflict recognition time, and defensive behavior. *Amer. Psychologist*, 1954, 9, 458-459.
- POSTMAN, L. The experimental analysis of motivational factors in perception. In M. R. Jones (Ed.), *Current theory and research in motivation*. Lincoln: Univer. Nebraska Press, 1953. Pp. 59-108.
- POSTMAN, L., BRONSON, WANDA C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, 48, 215-224.
- POSTMAN, L., BRUNER, J., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, 43, 142-154.
- ROSEN, C. C. Change in perceptual threshold. *J. Pers.*, 1954, 23, 182-194.
- ROSENZWEIG, S. The picture-association method and its application in a study of reactions to frustration. *J. Pers.*, 1945, 14, 3-23.
- SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, 43, 195-202.
- THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.
- ULLMANN, L. P. An empirically derived MMPI scale that measures facilitation-inhibition of recognition of threatening stimuli. *Palo Alto VA Hosp. res. Rep.*, 1960, No. 10.

(Received March 27, 1961)



## CRITIQUE AND NOTES

### CRIME AND PSYCHOPATHOLOGY

MALCOLM D. GYNTHYER

*Washington University Medical School*

MMPI studies of criminals have demonstrated that there are personality differences between this group and normals (Panton, 1959; Stanton, 1956). However, studies which have attempted to differentiate various types of criminal activity on the basis of MMPI data have met with little success (Levy, Southcombe, Cranor, & Freeman, 1952; Panton, 1958; Pothast, 1957).

One reason for this failure to differentiate among crime classification groups appears to be related to the common practice of discarding MMPI profiles with deviant validity scores. Stanton (1956), for example, discarded 89 out of 298 MMPI profiles because of their presumed invalidity; his criteria for elimination were: "raw scores of over 30 on the Question scale, 7 and over on the *L* scale and 17 and over on the *F* scale." A recent study (Gynther, 1961) has shown that so-called "invalid" MMPI *F* scores are meaningfully related to diagnosis and aggressive behavior and that elimination of such data prior to analysis leads to biased results, particularly in dealing with individuals who display antisocial actions.

Another reason for the lack of success in finding differences in personality and/or level of adjustment among crime groups seems to be associated with a failure to adjust for the effect of variables such as age, intelligence, and socioeconomic level which are known to influence MMPI profiles (Dahlstrom & Welsh, 1960). For example, Panton (1958) sorted 1,313 prisoners into six crime classification groups (i.e., aggravated assault, property theft, etc.), but gave no information concerning the above mentioned variables, except to state that prisoners with intelligence quotients below 80 had been excluded.

Another problem with MMPI research has to do with the extreme complexity of the data to be analyzed. In this connection, it would be helpful if a single MMPI score could be used as an indicator of adjustment or personality type. In their recent summary of MMPI research, Dahlstrom and Welsh (1960) state that "the *F* scale value probably serves as the best single measure of severity of illness." From the context it is clear that by "illness" the authors refer to emo-

tional disorders. Thus, a high *F* scale value would indicate severe psychopathology and a low *F* scale value would suggest that an emotional disturbance is mild or nonexistent.

The main purpose of this study is to examine the relationships between degree of psychopathology, as measured by the MMPI *F* scale, and type of criminal activity, with the effects of age, intelligence, socioeconomic status, sex, and race controlled. The relations obtained among the different variables will also be discussed.

#### METHOD

The subjects of this study were all the white male court referrals admitted to South Carolina State Hospital between September 1957 and August 1960 who completed the MMPI and Kent EGY, Scale D. One hundred and two court cases were unable or unwilling to take the MMPI due to illiteracy, mental deficiency, confusion associated with organicity, schizophrenic disorganization, or paranoid suspiciousness. The 251 cases which were available were divided into subgroups on the basis of the criminal activity with which they were charged. The 12 crime categories containing the most cases (cf. Table 1) were retained for the analysis, a procedure which resulted in the elimination of 61 cases with relatively infrequent charges against them (i.e., slander, involuntary manslaughter, etc.). The mean age of the remaining 190 cases was 32.52 years and the mean Kent score was 27.78 (the average range is 24 to 31 inclusive). More than 90% of this group were released from the hospital after one month's observation and later convicted and sent to prison by the courts.

The influence of sex and race on MMPI profiles was handled by restricting the experimental sample to white males. The effect of socioeconomic status is negligible due to the fact that the group was very homogeneous with respect to this variable. Consideration of the education, occupation, and income of each of these individuals prior to hospitalization indicated that nearly all of the 190 cases would be classified as *low* socioeconomic level. Adjustments for age and intelligence differences among the subgroups were made by analysis of covariance techniques.

#### RESULTS AND DISCUSSION

Table 1 presents the crime classes, number in each class, and means and standard deviations of



TABLE 1  
MEANS AND STANDARD DEVIATIONS OF THE MMPI *F*, AGE, AND KENT EGY  
VARIABLES FOR THE DIFFERENT CRIME CLASSES

Crime	<i>N</i>	MMPI <i>F</i> scores		Age		Kent scores	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Lewd acts on children	9	16.11	7.96	39.00	12.05	26.11	4.28
Rape, attempted rape	17	13.53	9.37	22.53	10.59	24.41	8.40
Incest	14	10.71	6.65	40.00	7.13	27.21	7.45
Larceny, stealing, theft	26	10.38	8.33	23.54	10.45	26.58	4.83
Murder	12	9.33	9.30	36.83	9.90	28.25	5.63
Forgery, passing bad checks	20	8.55	9.86	28.75	4.54	30.45	2.80
Assault and battery with intent to kill	15	6.20	4.71	44.33	15.15	29.66	4.12
Housebreaking and grand larceny	24	6.00	7.24	29.29	9.84	26.42	5.83
Nonsupport	10	4.90	2.73	36.40	9.18	27.80	5.33
Voyeurism, exhibitionism	11	4.36	2.94	28.54	9.88	28.45	5.03
Drunken driving	18	4.33	4.70	41.05	7.99	30.83	3.42
Assault and battery	14	4.28	4.58	33.64	11.53	27.43	4.83

the three variables for each class. The data have been arranged so that the largest mean MMPI *F* scores are at the top of the list and the smallest mean MMPI *F* scores are at the bottom. Examination of the table reveals wide differences among the crime groups with regard to the *F*, age, and Kent variables. It is noteworthy, for example, that the three highest mean *F* scores are obtained from groups charged with sexual crimes, a result which is consistent with Panton's (1958) finding that his Sexual Pervasive group gave the most deviant profile. On the other hand, the low mean *F* scores obtained from voyeurs and exhibitionists suggest that their inclusion in a Sexual Pervasive group with those who commit lewd acts against children and incest would tend to obscure important differences. It might be pointed out that the mean *F* score of the Lewd Acts group is so large that this group would be virtually eliminated if the usual practice of discarding  $F > 16$  profiles were followed.

Age appears related to type of offense. Stealing and rape appear to be crimes of youth, whereas assault and battery with intent to kill and drunken driving seem to be charges associated with early middle age. These results are similar to those reported by the FBI (Reckless, 1960), suggesting that our sample is representative of criminals-in-general. Of course, the very nature of some crimes (i.e., incest involving father and daughter) determines the possible range of ages. Intelligence, also, varies with offense, but in no very consistent manner. It is interesting that one of the so-called White Collar groups (i.e., forgers, bad check passers) obtained a high mean Kent score which is consistent with most statements about this crime category.

The major question raised by this study is whether crime and psychopathology are related, with other relevant variables controlled. Table 2 summarizes the analysis of covariance. The obtained *F* value of 3.30 is highly significant ( $p < .001$ ). This result indicates that the differences in mean MMPI *F* scale values among the crime groups cannot be accounted for by the differences among the groups in mean ages and mean Kent scores. If one accepts Dahlstrom and Welsh's (1960) assertion that the *F* scale is a measure of psychopathology, the conclusion is that crime groups obtaining large *F* values are emotionally "sicker" than crime groups obtaining small *F* values. In the present instance, then, one might say that sex criminals (excepting voyeurs and exhibitionists) are most emotionally disturbed and persons charged with assault and battery, drunken driving, voyeurism and/or exhibitionism, and nonsupport are least emotionally disturbed. It would be of interest to know whether our findings would also hold for convicted persons not referred for sanity examinations.

TABLE 2  
ANALYSIS OF COVARIANCE:  
MMPI *F* SCORE, AGE, AND KENT  
EGY SCORE

Source	<i>df</i>	<i>MS</i>	<i>F</i>
Adjusted means	11	171.76	3.30*
Within-groups	177	52.02	
Total	190		

\*  $p < .001$ .

It was stated earlier that most of these individuals were returned to the courts, whereas a minority were retained in the hospital.<sup>1</sup> The question may well be raised as to differences in *F* scores between these subgroups. One might hypothesize that only those with extremely high *F* scores would be retained in the hospital. Analysis of the data lends no support to this conjecture; the mean *F* score of those returned to the courts was 8.49 and the mean *F* score of those retained in the hospital was 7.50. There are variations in percentage of cases in each crime category retained in the hospital; however, rank-difference correlation between crime category as ranked by mean *F* score and as ranked by percentage of cases retained in the hospital is +.09 which indicates that these variations are not related to *F* score.

In view of our findings concerning persons who commit aggressive sexual acts, it is noteworthy that over 25 states and the District of Columbia have enacted special statutes, commonly referred to as sexual psychopath laws, which provide for commitment of the sex offender to a mental hospital (MacDonald, 1958). The offender is not released until he has recovered, or until he may be safely discharged without danger to society. In contrast to other criminal statutes, the special statutes do not require that the offender be legally insane to qualify for treatment in a mental hospital rather than detention in a prison. The existence of a legal mechanism for handling such cases suggests that society is aware that sexual offenders are people with severe emotional disorders.

<sup>1</sup> A table representing MMPI *F* scores of court cases sent to prison versus those retained in hospital has been deposited with the American Documentation Institute. Order Document No. 7060 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

## SUMMARY

This study investigated the relations between type of crime, age, intelligence, and degree of psychopathology, as measured by the MMPI *F* score. MMPIs were available from 190 white male court referrals who were nearly all classified as low socioeconomic status. It should be noted that one crime class (i.e., Lewd acts on children) would have been excluded from the analysis if the usual criteria for discarding MMPIs had been used.

A definite relationship between type of crime and degree of psychopathology was found. Individuals who commit sexual crimes appear to be the most seriously emotionally disturbed. No difference in mean *F* score between subjects released to the courts versus those retained in the hospital was found.

## REFERENCES

- DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook*. Minneapolis: Univer. Minnesota Press, 1960.
- GYNTHIER, M. D. The clinical utility of "invalid" MMPI *F* scores. *J. consult. Psychol.*, 1961, 25, 540-542.
- LEVY, S., SOUTHCOMBE, R. H., CRANOR, J. R., & FREEMAN, R. A. The outstanding personality factors among the population of a state penitentiary: A preliminary report. *J. clin. exp. Psychopath.*, 1952, 13, 117-130.
- MACDONALD, J. M. *Psychiatry and the criminal*. Springfield, Ill.: Charles C Thomas, 1958.
- PANTON, J. H. MMPI profile configurations among crime classification groups. *J. clin. Psychol.*, 1958, 14, 305-308.
- PANTON, J. H. The response of prison inmates to seven new MMPI scales. *J. clin. Psychol.*, 1959, 15, 196-197.
- POTHAST, M. D. A personality study of two types of murderers. *Dissertation Abstr.*, 1957, 17, 898-899.
- RECKLESS, W. R. *The crime problem*. (3rd ed.) New York: Appleton-Century-Crofts, 1960.
- STANTON, J. M. Group personality profile related to aspects of antisocial behavior. *J. crim. Law Criminol. police Sci.*, 1956, 47, 340-349.

(Received March 24, 1961)

## EGO IMPAIRMENT IN SCHIZOPHRENIA AS REFLECTED IN THE OBJECT SORTING TEST

LLOYD H. SILVERMAN<sup>1</sup>

*Jewish Board of Guardians and New York University*

AND DORIS K. SILVERMAN

*New York University*

Beres (1956) and Bellak (1958), utilizing a psychoanalytic framework, conceptualize the psychopathology involved in schizophrenia in terms of ego disturbance. Considering the ego as a substructure of personality defined by its functions, these authors list the following ego functions as being severely impaired in this disorder: relations to reality, regulation and control of instinctual drives, object relations, thought processes, the defensive functions, the autonomous functions, and the synthetic function.

In their discussion of *thought processes*, these authors characterize the schizophrenic with such terms as "autistic," "loose," "prelogical," and "irrational." The Object Sorting Test lends itself to an examination of these characteristics of thought in general and concept formation in particular as evidenced by the studies of McGaughran and Moran (1956a, 1957) and Leventhal, McGaughran, and Moran (1959). Consistent with Beres' (1956) and Bellak's (1958) descriptions of schizophrenic thinking, McGaughran and Moran speak of the concept formation of these patients as involving "idiosyncratic forms of symbolism, overabstractions, and . . . ellipses in the conceptual process" and as based on "a synechdochal, or other type of loose-representational relationship" (p. 48). The term they use to describe these typically schizophrenic conceptualizations is "open-private," private referring to the conceptualization not being one "shared and freely communicated by the majority of persons within the same culture" and open referring to the principle being insufficiently delimiting. These authors have demonstrated that open-private conceptualizations on the Object Sorting Test occur significantly more often in schizophrenics than in brain damaged patients (Leventhal et al., 1959; McGaughran & Moran, 1957) and nonpsychiatric patients (McGaughran & Moran, 1956a). However, a survey of the literature reveals that no study has demonstrated that schizophrenics produce more open-private responses than patients suffering from other kinds

of psychiatric disturbance.<sup>2</sup> Such studies are needed if one wishes to conclude that this type of conceptualization is specifically characteristic of schizophrenia rather than being characteristic of psychiatric disturbance in general. Thus, one of our purposes in the present research is to investigate if this conclusion can be drawn. Our second purpose, derived from clinical leads (Bellak, 1958; Beres, 1956), is to investigate other aspects of disturbed ego functioning on the Object Sorting Test which should distinguish schizophrenics from patients with other kinds of psychiatric disorders.

### METHOD

#### *Subjects*

Subjects were 40 adolescents (ages 11-18) who were patients at the Hawthorne Cedar Knolls and Linden Hill schools, open residential treatment centers that are divisions of the Jewish Board of Guardians. They were divided into two groups, one consisting of 20 schizophrenics and the second of 20 psychoneurotics and personality disorders (as these terms are used in the diagnostic manual of the American Psychiatric Association, 1952). All 40 subjects were patients for whom there was concurrence on diagnosis by at least two clinicians, usually the psychiatric social worker who was treating the patient and a psychiatrist. The subjects were matched in pairs for IQ, age, and sex. There was no significant difference between the mean ages for the schizophrenic and neurotic-personality disorder groups (15.7 and 15.8, respectively) or between the mean IQ (105.8 and 103.4). There were 17 males and 3 females in each group.

<sup>2</sup>A possible exception to this statement are the findings of Rapaport (1945). He compared a group of schizophrenics with other psychiatric groups for some of the kinds of responses McGaughran and Moran include in their open-private category, (though referring to them by different names), and offers some empirical support for the differentiating power of a few of these signs (p. 452). However, even for these, the results are badly in need of further validation since: the differentiating ability of the signs was arrived at ad hoc; tests of significance were not reported; no indications were given that the groups involved were equated for age or IQ; it is not clear whether a "blind" evaluation of the data was undertaken.

<sup>1</sup>Now at University of Pennsylvania, School of Medicine.



### Procedure

The Rapaport (1945) modification of the Goldstein-Gelb-Weigl Object Sorting Test was administered individually to each subject according to the procedure described by Rapaport. Both writers took part in the administration, with each of us testing an equal number of patients in each group. One of us (DKS), who had no knowledge of the subjects' clinical diagnoses, "blindly" rated each protocol, placing responses into the following four categories:

1. The Open-Private category of McGaughran and Moran (1956b). These authors have delineated 11 types of open-private responses which they describe in their scoring manual.

2. The Essential Looseness category described by Rapaport (1945, p. 406). We include this category for two reasons. First, it seems to involve aspects of the kind of thinking that was described in Category 1, only it shows itself in the number of items included in the sorting rather than in the verbalization. Second, it was one of the four categories that Rapaport found best differentiated schizophrenic groups from both other clinical groups and a nonpsychiatric patient group (p. 452).<sup>3</sup>

3. Intrusion of Instinctual Drives. We included here responses in which the content of the subjects' verbalization, or his behavior while giving the response, revealed an inappropriate expression of an instinctual drive. We felt that such responses should be scored since they reflect an impairment in one's ability to control such drives which Beres (1956) and Bellak (1958) both have described as characterizing the schizophrenic. For subcategorizing these responses, we used the breakdown employed by Holt (1960) in his "Manual for the Scoring of Primary Process Manifestations in Rorschach Responses." Thus, scores were given for responses reflecting drives with libidinal aim (that is, responses with oral, anal, sexual, homosexual, or exhibitionistic-voyeuristic implications) and aggressive aims. The following are some of the responses so scored: (a) Libidinal—a response referring to the imitation cigarette as being "like candy" was scored for its oral implications. (However, if the subject stated that the silverware are "objects you eat with," no score was given. The objects involved here legitimately pertain to eating and, thus, no intrusion of an instinctual drive could be said to have taken place.) A response of smelling the pipe on the smoking equipment item, was scored for its anal implications. (b) Aggressive—a score was given for grouping the lock with the bell and stating "this is a bell for warning. Let's say the burglar is trying to get in a store. This lock would be on. When he

<sup>3</sup> The other three categories that Rapaport (1945) alludes to in this context are "symbolic," "fabulated," and "syncretistic," which refer to the subject's verbalization. Most of the responses falling under these headings are subsumed under McGaughran and Moran's open-private category, though described by other terms.

breaks it, the bell goes off." A score was given for a response to the white objects (cigarettes, sugar, and filing card) by stating "all can be destroyed very easily."

4. Miscellaneous. This category was established to cover aspects of the subject's verbal or nonverbal reactions during testing that, in the "blind" rater's judgment, reflected aspects of schizophrenic ego disturbance that could not be scored in any of the three prior categories. After all the protocols were scored and the data tabulated, the Miscellaneous responses were re-examined. It was found that a preponderance of them could be grouped into four subcategories, two of which primarily seemed to reflect an impairment in thought processes (Subcategories a and b) and two of which mainly appeared to reveal a disturbance in the patient's relation to reality (Subcategories c and d). These ad hoc groupings are presented below together with the kinds of responses that were scored in each of them: (a) Peculiar Verbalizations—a designation used by Holt (1960) in his Rorschach manual to describe "instances of linguistic usage that are autistic enough to sound odd, recognizably not idiomatic, and not a function of group membership, unfamiliarity with English, or the like" (p. 35).<sup>4</sup> For example, one subject referred to the toys as "all small images of their main purpose." (b) Perseverative Thinking—refers to the inappropriate use of a conceptualization that had been used appropriately on an earlier item. (c) Inappropriate Motor Activity—the subject behaved in the test situation inappropriately, using the objects in a manner that violated the explicit and implicit restrictions imposed by the test instructions. For example, one subject started to remove a screw from the bell with a toy screwdriver and seeing that it could be done decided that the objects belonged together. (d) Out of Field Responses—the subject introduced the idea of including in a sort objects that were not part of the test. Thus, one subject chose the lock as his free choice item and then pointed to the door of the testing room and stated, "they [usually] put a lock [on the door]."

Since it was possible for a response to have different scorable elements, it could be scored more than once. Most elements that were scored were given weights of one. However, if the rater felt that the element was an extreme example of pathology in a particular category, she gave it double weight. If she felt that it was questionable whether the element fell into the category, she gave it half weight. Thus, a three-point scale was available (.5, 1, and 2).

After scoring the responses of a protocol, the "blind" rater made an overall diagnostic appraisal. She used as her criterion her global impression of the protocol based on a flexible weighing of the responses in the four categories discussed above. In making these judgments, each protocol was classified in one of four ways: schizophrenic; psychoneurotic

<sup>4</sup> Schafer (1948), in his individual case studies, uses this term much as we do to describe certain aspects of Object Sorting Test responses.

TABLE 1

MEAN SCORES FOR THE TWO PSYCHIATRIC GROUPS AND *t* TESTS FOR DIFFERENCES BETWEEN THEM

Category	Schizophrenics		Neurotics and personality disorders		<i>t</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Open-Private	5.25 (5.2)	3.84	3.33 (3.6)	3.06	1.76*
Essential looseness	.73	1.37	.63	1.06	.26
Intrusion of instinctual drives	1.10	1.49	.25	.41	2.46***
Miscellaneous	2.05	2.06	.95	1.11	2.08**
Total	9.13	7.27	5.15	4.25	2.10**

\* Significant at .05 level.

\*\* Significant at .025 level.

\*\*\* Significant at .01 level.

or personality disorder; don't know, but a forced choice would be schizophrenic; don't know, but a forced choice would be psychoneurotic or personality disorder.

### Hypotheses

1. The total scores for the four categories will be significantly greater for the schizophrenics than for the psychoneurotics and personality disorders.
2. When the protocols are evaluated globally the two diagnostic types can be distinguished from each other to a degree greater than that expected on the basis of chance.

### RESULTS

The mean scores for the four categories<sup>5</sup> are presented in Table 1.

The results bearing on Hypothesis 1 can be found in the last row which presents the means for all four categories combined. The difference

<sup>5</sup> For the open-private category, the numbers in parentheses are the mean scores for the two groups, when each scored response is given an equal weight rather than being placed on a three-point scale. These means are included since we felt it would be desirable to compare the findings on our schizophrenics with those of McGaughran and Moran (1956a) and Leventhal, McGaughran, and Moran (1959) who did not differentially weight the scores they assigned.

between the two groups is significant at the .025 level (one-tailed *t* test).<sup>6</sup> When each of the four categories is considered separately, the differences are all in the predicted direction, and for all but Category 2, these differences are significant below the .05 level. Thus, Hypothesis 1 is strongly supported by the findings of this study.

The results bearing on Hypothesis 2 are presented in Table 2.

When all the global judgments are considered, 22 of them were correct and 18 incorrect. These results are not significant (binomial test). When the "don't know but forced choice" judgments are excluded, 19 cases remain in which the rater felt reasonably sure of her judgment. She was correct in 13 or 68% of these. Although the difference is not significant, there is a tendency in the predicted direction ( $p = .08$ , one-tailed binomial test). However, Hypothesis 2 cannot be said to be more than weakly supported by these results.

<sup>6</sup> The numbers presented are combined totals for Part I and Part II of the Object Sorting Test since the two parts were almost equally effective in differentiating the two psychiatric groups. Of the scores assigned for Part I, 64.6% of them were given to the schizophrenics compared to 63.4% for Part II.

TABLE 2  
DIAGNOSTIC JUDGMENTS

Actual status	Judgment of status				
	Schizophrenic	Don't know schizophrenic	Total schizophrenic	Neurotics and personality disorders	Total Neurotic and Personality Disorders
Schizophrenics	5	6	11	4	9
Neurotics and personality disorders	2	7	9	8	11



## DISCUSSION

The results of this study take on particular significance when they are considered in conjunction with the findings of McGaughran and Moran. These authors demonstrated that schizophrenic adults could be differentiated from brain damaged and nonpsychiatric patient adults on the sorting test for those responses they called open-private. The results of the current study allow for the following additional conclusions to be drawn:

1. These open-private responses not only can differentiate schizophrenics from organics and nonpsychiatric patients, but they can differentiate them from psychoneurotics and personality disorders. This is an important finding for the clinical use of the sorting test for differential diagnosis. What makes the findings in this investigation particularly noteworthy is that our schizophrenics did not require hospitalization, and, thus, do not represent the "sickest" segment of the schizophrenic population while the nonschizophrenics *did* require residential treatment, suggesting they have greater ego disturbance than most patients with similar diagnoses. Thus, one would have expected the usual differences in ego functioning between schizophrenics, on the one hand, and neurotics and personality disorders, on the other, to have been reduced by the particular samples of patients we used here. The fact that there were still significant differences between the groups therefore attests to the differentiating power of McGaughran and Moran's open-private category.

2. These open-private responses characterize adolescent schizophrenics as well as adult schizophrenics. The average number of such responses for our adolescent group was 5.2 which is close to the 5.7 average for one adult schizophrenic group studied (McGaughran & Moran, 1956a). It was a good deal lower than the 9.5 average for a second adult schizophrenic group (Leventhal et al., 1959). The difference between the two adult groups in their average open-private responses has been ascribed to the fact that the second group was hospitalized for a greater length of time.

The greater degree of chronicity for the [second] group, then, may be associated with a greater tendency to employ open-private sortings, a type of conceptualization which presumably would increase with increased social withdrawal (p. 89).

Since our adolescent schizophrenic population was a nonhospitalized group, it is understandable that their average is closer to that of the first adult schizophrenic group than the second.

3. While the open-private category covers many of the responses that reflect aspects of the ego impairment in schizophrenia, there are other responses, not covered by this category which are at least as characteristic. We saw these responses as reflecting impairment in three ego functions, which according to the psychoanalytic position expounded by Beres (1956) and Bellak (1958), are among those functions most severely disturbed in schizophrenia. These responses can be grouped into those primarily reflecting a disturbance in the thought processes (including Peculiar Verbalizations and Perseverative Thinking), relation to reality (including Inappropriate Motor Activity and Out of Field responses), and regulation and control of instinctual drives. Subject to confirmation by other investigators, we tentatively conclude that the use of a scoring system (for both research and clinical practice) that considers these responses in addition to those designated as open-private<sup>7</sup> will result in a sharper differentiation between schizophrenics and patients with other kinds of psychiatric disturbances.

## SUMMARY

A group of 20 schizophrenic adolescents were compared with a group of 20 psychoneurotics and personality disorders for their performance on the Object Sorting Test. The subjects in the two groups were matched for age, IQ, and sex, and all were patients in an open residential treatment center. The test protocols were "blindly" evaluated for four categories of responses for which it was hypothesized that the schizophrenics would have higher scores. It was also hypothesized that the rater, making a global judgment as to each patient's diagnosis, could significantly differentiate between the two groups.

While the second hypothesis was only weakly supported by the results, the first hypothesis was strongly supported. Significant differences in the predicted direction were found both overall and for three of the four categories studied. These were: McGaughran and Moran's "open-private" responses; responses reflecting inappropriate intrusions of instinctual drives; a group of miscellaneous responses including those labeled Peculiar Verbalizations, Perseverative Thinking, Inappropriate Activity, and Out of Field. These findings were seen as reflecting different aspects of the ego impairment in schizophrenia.

<sup>7</sup> The open-private responses seem to use to involve a disturbance in two of the aforementioned ego functions, thought processes and relations to reality.



## REFERENCES

- AMERICAN PSYCHIATRIC ASSOCIATION, Mental Hospital Service, Committee on Nomenclature and Statistics of the American Psychiatric Association. *Diagnostic and statistical manual: Mental disorders*. Washington, D. C.: APA, 1952.
- BELLAK, L. (Ed.) *Schizophrenia: A review of the syndrome*. New York: Logos, 1958. Ch. 1.
- BERES, D. Ego deviation and the concept of schizophrenia. *Psychoanal. Stud. Child.*, 1956, 11, 164-235.
- HOLT, R. R. Manual for the scoring of primary process manifestations in Rorschach responses. (Draft 8) New York University, Research Center for Mental Health, 1960. (Ditto)
- LEVENTHAL, D. B., MCGAUGHRAN, L. S., & MORAN, L. J. Multivariable analysis of the conceptual behavior of schizophrenic and brain-damaged patients. *J. abnorm. soc. Psychol.*, 1959, 58, 84-90.
- MCGAUGHRAN, L. S., & MORAN, J. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and nonpsychiatric groups. *J. abnorm. soc. Psychol.*, 1956, 52, 43-58. (a)
- MCGAUGHRAN, L. S., & MORAN, L. J. Revised (1956) supplementary instructions for scoring conceptual area on the Goldstein-Gelb-Weigl Object Sorting Test. 1956. (Ditto, Author, University of Houston) (b)
- MCGAUGHRAN, L. S., & MORAN, L. J. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object sorting. *J. abnorm. soc. Psychol.*, 1957, 54, 44-49.
- RAPAPORT, D. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.
- SCHAFER, R. *Clinical application of psychological tests*. New York: International Univer. Press, 1948.

(Received April 3, 1961)

A RELATIONSHIP BETWEEN SELF-ESTEEM AND PERSUASIBILITY<sup>1</sup>

HOWARD LEVENTHAL AND SIDNEY I. PERLOE<sup>2</sup>

Yale University

One goal of communications research has been to isolate variables which can account for individual differences in the acceptance of persuasive communications. The self-esteem or favorableness of self-rating of the recipient is one such variable. It has been suggested that persons low in esteem tend to see themselves as different and less worthy than others and are therefore more accepting of influence (Lesser & Abelson, 1959). Relationships between self-esteem and persuasibility, however, have often not been very high (Janis, 1954, 1955; Janis & Field, 1959) or have been found only in restricted groups (Janis & Field, 1959) and under special circumstances (Lesser & Abelson, 1959).

A somewhat different approach to self-esteem has been taken by Cohen (1959). He has proposed that people who rate themselves high in esteem use avoidance defenses, while those who rate themselves low incline towards expressive or sensitizing defenses. Evidence for this relationship has been presented by Cohen (1959, p.

115) and by Coopersmith (1959). On the basis of this view of self-esteem, Cohen has hypothesized that "threatening appeals may be rejected more by those of high self-esteem than those of low self-esteem. On the other hand, appeals which enhance an individual's self picture might be accepted more by the highs than by the lows" (p. 119). A recent experiment by Goldstein (1959) showing that avoiders change more in response to minimal than to high fear appeals also provides partial support for the hypothesis.

The study to be reported has produced additional data which are relevant to Cohen's (1959) hypothesis concerning the relationship between esteem and persuasibility. Although the experiment was designed for another purpose, it did yield data relevant to the relationship between esteem and persuasibility. The research was originally constructed to explore the effects of similarity in personality characteristics of communicator and recipient upon attitude change. The specific hypothesis tested was that subjects who receive communications opposing their own positions from communicators who are similar to them on relevant personality characteristics would show more attitude change toward the communication than subjects receiving opposing communications from communicators who are dissimilar to them. The data did not support the hypothesis, although, as will be shown below, similarity did have an effect on attitude change.

<sup>1</sup> Supported in part by Mental Health Grant M-3575 to Leventhal and NSF Grant G13023 to Perloe. The authors wish to express their appreciation to the late Carl Hovland of Yale University for his helpful criticism of an earlier form of the present paper. They also wish to thank G. Whitney Biggs for his help in aspects of the planning and running of the study and to Alan Chasnoff for his help in analyzing the data.

<sup>2</sup> Now at Haverford College.

The remainder of the paper will focus on Cohen's (1959) hypothesis and will present only those aspects of the original method and results which are relevant to this hypothesis.

## METHOD

### Study Design

*Attitude content.* The issue selected was attitude toward Army life. This included many subtopics such as reactions to authority, feelings concerning an all male environment, conflicts with independence and intellectual needs, etc. The issue involves many potential threats while at the same time it is of great interest to students.

*Communications.* Subjects received what purported to be interview transcripts of ex-college men who were in the Army. The communications were presented supposedly as a stimulus for a discussion of Army life. Because initial attitudes toward the Army were equally divided in favorability, two persuasive appeals were prepared. In both, the communicator had been unexpectedly drafted and had had his career plans interrupted. The individual summarized his feelings toward Army life as they developed during his service. While the favorable communication began with the soldier expressing his initial doubts and annoyance with Army life, his reaction became progressively more favorable. The communication might be characterized as optimistic and buoyant, with the soldier making a highly adequate adjustment to the Army. In each situation he looked at the "bright side" and secured many gratifications for himself. The anti-Army communication was negativistic, angry, and clearly showed an individual engaged in trivial, boring, and self-demeaning activities. Subjects with initially positive attitudes received the anti-Army communication; those with negative attitudes received the pro-Army communication.

Each communication covered identical topics: basic training, the typical Army sergeant, overseas duty, job placement and work experience, friendships in service, sex activities, and the degree to which the soldier felt he benefitted from his entire Army experience. They were comparable in expressiveness and verbal skill.

*Similarity.* Before reading the interview transcripts, subjects were given personality ratings of the man whose comments they were about to read. They were told that knowing about the man would help them in evaluating his reactions. A special set of ratings was constructed for each subject. In one condition these ratings were quite similar to ratings the subject had previously made of himself; in the other, they were quite dissimilar.

*Esteem.* Self-esteem was introduced in the original study for two purposes. First, it was to serve as a control variable for the similarity manipulation. There was a danger that dissimilar communicators for high esteem subjects would be low in esteem, and those for low esteem subjects would be high. To control for these effects, each communicator and recipient were matched in esteem. Second, inclusion of the esteem variable permitted an examination of the relationship between self-esteem and persuasibility.

Thus the design involved two levels of three variables: similarity, esteem, and favorableness of communication.

### Initial Measures

*Initial attitudes.* The initial measures were collected at least one week prior to the attitude change session. Subjects rated Army life on 22 characteristics. The applicability of each adjective and phrase to Army life was indicated on a seven-point scale. Desirability ratings for each characteristic were also obtained on seven-point scales. The applicability and desirability ratings were scaled from +3 (highly applicable and highly desirable) to -3 (inapplicable and undesirable). An overall index of attitude toward Army life was obtained by multiplying the applicability and desirability ratings and summing across the 22 characteristics. The sample was split into favorable and unfavorable groups at the zero point, with favorable attitudes above zero and unfavorable below. Had the split been made at the median, which was 1.5, only one subject would have been reclassified.

*Self-esteem.* Subjects rated themselves on 27 characteristics. Twenty-two of the adjectives were descriptive of personality traits and five indicated an interest in music, art, science, etc. Ratings were on seven-point scales. Twenty of these traits and interests were rated for desirability on seven-point scales. Both the applicability and desirability ratings for the 20 characteristics were converted into scale values (+3 to -3 as in the attitude questionnaire), multiplied, and summed. The score provides a measure of self-esteem, or more accurately a measure of the degree to which a person attributes positively valued characteristics to himself and disavows negative characteristics. Subjects were divided into high and low self-esteem groups using a median split.

### Subjects

A total of 58 male undergraduates from introductory courses in psychology and sociology were used in the present study. None of the subjects had served in the Army or were in the ROTC.

### Procedure

Subjects were seen in groups, usually varying in size from four to six members; only three of the groups had fewer than four people and one had seven. To minimize suspicion concerning the communications and attitude measures, the sessions were carefully planned to emphasize their avowed purpose as discussions of Army life. The aim of the discussion, as described to the subjects, was to obtain "considered opinions and ideas" to amplify the previously obtained questionnaire material. It was first stated, however, that each participant would be given information about the personality of one particular soldier and his reaction to Army life.

Following the introductory instructions, each subject was given a Summary Rating Sheet with the ratings of the communicator's personality. These forms, presumably from the Army, were multilined, used several type sizes, and borders, special form numbers, etc., and looked very different from the



other questionnaire materials. After examining the forms, subjects rated the clarity of their impression, perceived similarity to the source, and their expectations that his reaction to Army life would clarify their own feelings. The communications were next given out on special forms marked Interview Continuation Forms, which were also multilithed. While handing out and collecting the Summary Rating Sheets and communications, the experimenters made a variety of "spontaneous" remarks urging care in handling, since they had to be returned to the government.

Following the communication, it was stated that, "to maximize the meaningfulness and significance of the discussion we would like you to review for yourselves your own attitude toward Army life." Since it was necessary for all subjects to review their opinions in a like manner, each completed once again the Army attitude questionnaire. The investigator stressed the point that the procedures were necessary to add depth to the discussion and to avoid superficial debate. While one investigator was passing out the attitude questionnaires, the other was busily turning on and "checking" a tape recorder which was kept in a prominent position at all times. Identification numbers were also placed at each subject's position. A 15-minute discussion was actually held and recorded after which subjects were disabused of the idea that the materials were real and answered some questions on what they thought the experiment was about. They were asked not to communicate about the proceedings to others and were told the results would be given to them shortly.

## RESULTS

### Manipulation Checks

**Communications.** Differences were computed between the subject's attitudes on the premeasures and his attitudes following the communication about Army life. These data are presented in Table 1. An analysis of variance revealed significant differences between groups receiving positive and negative communications ( $F = 96.14$ ,  $df = 1/50$ ,  $p < .001$ ). The communications clearly had opposite effects on the subject's attitudes though the absolute amount of change generated by each was not significantly different.

**Similarity.** The manipulation of similarity created differences in the ratings (on a nine-point scale) of perceived similarity; means for the two conditions were 3.74 for the similar communicator, and 5.45 for the dissimilar communicator ( $F = 13.10$ ,  $df = 1/54$ ). Low numbers indicate greater similarity. There were no significant effects of esteem upon perceived similarity.

**Insight into experiment.** Checks were made on the degree to which subjects seemed to understand the purpose of the experiment. Subjects' free responses were scored for insight and there were no significant differences between conditions on this factor. Nearly all subjects thought the

TABLE 1  
MEAN CHANGE IN ATTITUDE FROM  
PRE- TO POSTQUESTIONNAIRES

Condition	Similar communicator		Dissimilar communicator	
	High Esteem	Low Esteem	High Esteem	Low Esteem
Received Pro-communication	+17.29 <sup>a</sup> (7) <sup>b</sup>	+18.00 (5)	+36.17 (6)	+5.10 (10)
Received Anti-communication	-21.38 (8)	-20.86 (7)	-9.50 (8)	-20.71 (7)

<sup>a</sup> A positive value indicates a change toward a more favorable attitude concerning Army life; a negative value indicates a change toward a less favorable attitude. The absolute amount of change is indicated by the magnitude of the value, disregarding sign.

<sup>b</sup> Number cases per cell.

Army was interested in college students' attitudes toward Army life in order to eventually improve these attitudes.

### Esteem and Similarity

An inspection of Table 1 reveals that high and low esteem subjects differed in their responsiveness to the two persuasive appeals. These differences in persuasibility were restricted, however, to the dissimilar communicator condition. Under this condition, high esteem subjects tended to change more toward optimistic communications than pessimistic ones, while low esteem subjects showed the opposite tendency. The test of this interaction was significant ( $F = 8.20$ ,  $df = 1/50$ ,  $p < .01$ ). The difference in the absolute amounts of attitude change shown by high esteem subjects to the two communications was significant ( $t = 2.48$ ,  $p < .025$ , two-tailed test) as was the difference between the absolute amount of change among the low esteem subjects ( $t = 2.05$ ,  $p < .05$ , two-tailed test).

A correlation was computed between self-esteem and initial attitude to see whether the above finding relating esteem and attitude change could have been due to differences in initial attitudes toward the Army at different esteem levels. The negligible correlation,  $r = -.07$ , seems to rule out such a possibility.

## DISCUSSION

The relationship between self-esteem and attitude change in the dissimilar condition seems to support Cohen's (1959) interaction hypothesis, rather than a generalized form of the simpler hypothesis of an inverse linear relationship between esteem and persuasibility (Janis, 1954; Janis & Field, 1959).

The high self-esteem subjects who have been assumed to be avoidance oriented were more



readily influenced by optimistic, gratifying, potentially self-enhancing communications than by pessimistic, threatening ones. Low esteem subjects who have been assumed to use expressive or sensitizing defenses showed the opposite effect.

There seem to be two factors responsible for the support given to Cohen's (1959) hypothesis. First, the issue of Army life posed a threat to the subjects' (Yale students) conceptions of themselves and of the type of life they believed they ought to lead. Second, the communications provided modes of anxiety reduction which varied in their compatibility with the subjects' defensive styles. The negative communication dwelt upon the anxiety laden content. It thereby stimulated preparatory worry (Janis, 1958), which was quite consistent with the defenses presumed to characterize the low esteem subjects. Such defenses reduce the subject's anxiety either by projecting or displacing his concern with the actual source of his anxiety to one which is less threatening, or by increasing his vigilance for future real threats and thereby reducing the possibility of his being overwhelmed by surprise. The positive communication, on the other hand, provided a useful means of denying the threat posed by Army life and was therefore consistent with the presumed high esteem defensive orientation. In either case subjects were more likely to be influenced by messages which they could use in their habitual modes of anxiety reduction, if these communications were provided by dissimilar communicators.

Neither the theory nor the data provide a clear reason for the relationship discussed above being limited to messages stemming from dissimilar communicators. One plausible explanation holds that the rejection of a communication from a similar source is akin to rejection of the self. This may be particularly important in the present case in which the message describes the communicator's personal reactions. A defensive rejection of the message under these circumstances might arouse more anxiety than if the defenses were not permitted to operate. With messages from a dissimilar source, this obstacle to the operation of the subjects' defenses is not present and the tendencies described above are able to affect persuasibility.

It is interesting to compare this experiment with Goldstein's (1959). He found a significant interaction between defense preference and level of fear appeal. Assuming the correspondence between high esteem and avoidance, and sensitization and low esteem, our findings agree on the former defensive orientation, but not on the latter. In his results, avoiders showed significantly more change to minimal than high fear appeals;

however, sensitizers changed equally to both messages. The difference in findings is probably due to differences in the communications. While Goldstein's messages differed in the degree to which they promoted anxiety, both were fear provoking. Each may have exceeded the threshold to promote "worry" by the sensitizer, and neither was of such positive and gratifying tone as to actually interfere with the sensitizer's propensity to ruminate on his worries.

#### SUMMARY

It has been proposed that people high in self-esteem tend to use avoidance defense mechanisms which lead them to reject threatening persuasive communications and to be receptive to optimistic messages. On the other hand, people low in self-esteem tend to use expressive or sensitizing defenses which lead them to reject optimistic appeals and accept threatening ones. Relevant data were presented from a study designed for another purpose.

It was found that subjects high in self-esteem were influenced more by optimistic communications than by threatening communications while subjects low in self-esteem showed the opposite pattern, but these results occurred only among subjects who received communications from sources dissimilar to subjects with respect to personality characteristics.

#### REFERENCES

- COHEN, A. R. Some implications of self-esteem for social influence. In C. I. Hovland & I. L. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale Univer. Press, 1959. Pp. 102-120.
- COPPERSMITH, S. A method for determining types of self-esteem. *J. abnorm. soc. Psychol.*, 1959, 59, 87-94.
- GOLDSTEIN, M. J. The relationship between coping and avoiding behavior and response to fear arousing propaganda. *J. abnorm. soc. Psychol.*, 1959, 58, 247-252.
- JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, 22, 504-518.
- JANIS, I. L. Anxiety indices related to susceptibility to persuasion. *J. abnorm. soc. Psychol.*, 1955, 51, 663-667.
- JANIS, I. L. *Psychological stress*. New York: Wiley, 1958.
- JANIS, I. L., & FIELD, P. B. A behavioral assessment of persuasibility: Consistency of individual differences. In C. I. Hovland & I. L. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale Univer. Press, 1959. Pp. 29-54.
- LESSER, G. S., & ABELSON, R. P. Personality correlates of persuasibility in children. In C. I. Hovland & I. L. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale Univer. Press, 1959. Pp. 167-186.

(Received April 18, 1961)

# Journal of Abnormal and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*

Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

## CONSULTING EDITORS

JOSEPH B. ADELSON  
*University of Michigan*

GORDON W. ALLPORT  
*Harvard University*

NORMAN H. ANDERSON  
*University of California, Los Angeles*

LEONARD BERKOWITZ  
*University of Wisconsin*

JACK W. BREHM  
*Duke University*

URIE BRONFENBRENNER  
*Cornell University*

DONALD R. BROWN  
*Bryn Mawr College*

DONALD T. CAMPBELL  
*Northwestern University*

LAUNOR F. CARTER  
*System Development Corporation*

DWIGHT W. CHAPMAN, JR.  
*Vassar College*

MORTON DEUTSCH  
*Bell Telephone Laboratories*

SEYMOUR FESHBACH  
*University of Pennsylvania*

DONALD W. FISKE  
*University of Chicago*

NATHANIEL L. GAGE  
*University of Illinois*

NICHOLAS HOBBS  
*Peabody College for Teachers*

MURRAY HORWITZ  
*New York University*

IRVING L. JANIS  
*Yale University*

HAROLD H. KELLEY  
*University of California, Los Angeles*

HERBERT C. KELMAN  
*Harvard University*

DANIEL J. LEVINSON  
*Harvard Medical School*

DAVID C. MCCLELLAND  
*Harvard University*

ROBERT B. MACLEOD  
*Cornell University*

HAROLD L. RAUSH  
*National Institute of Mental Health*

NEVITT SANFORD  
*Stanford University*

IRVING SARNOFF  
*Western Reserve University*

WILLIAM A. SCOTT  
*University of Colorado*

JANET TAYLOR SPENCE  
*Veterans Administration Hospital,  
Iowa City*

JOSEPH VEROFF  
*University of Michigan*

JUSTIN L. WEISS  
*Massachusetts Mental Health Center*

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued monthly, two volumes per year. The subscription price per volume is \$10.00, or \$20.00 annually (foreign, \$20.50). Single issues are \$2.00.

Original manuscripts should be sent to the Editor:

DANIEL KATZ  
University of Michigan  
Box 1248  
Ann Arbor, Michigan

---

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR  
Acting Production Manager

JUDY SMITH  
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
1333 Sixteenth Street, N.W.  
Washington 6, D. C.

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.  
Prince and Lemon Sts., Lancaster, Pa.  
and 1333 Sixteenth Street N.W., Washington 6, D. C.  
Second-class postage paid at Lancaster, Pa., and at additional mailing offices  
Copyright © 1962 by the American Psychological Association, Inc.



# COMMUNALITY AND STABILITY OF MEANING IN CLINICAL CASE DESCRIPTION<sup>1</sup>

ERIKA CHANCE, JACK ARNOLD, AND SYBIL TYRRELL

*Department of Psychiatry, Mount Zion Hospital, San Francisco*

One major difficulty in communication among clinicians is that much depends on the point of view of the observer and on his peculiar emphasis of selected aspects of behavior as crucial variables in case description. Most clinical investigations employ variables for describing behavior without questioning whether the variables mean the same thing when applied by different observers or when applied to different patients. In many such studies it is commonly assumed that each of the variables selected corresponds to a unitary concept that is used with more or less emphasis by various kinds of observers (observer bias) and is found to be more or less characteristic of various kinds of patients (differences among patients). These studies focus on differences in score level among observer or patient populations.

In our earlier work (Chance, 1959; Chance & Arnold, 1960) with a system of variables for classifying interpersonal experiences (see Figure 1), we assumed, first, that the 20 variables which compose the system had the same meanings for a variety of observers and that the meaning of each variable was stable regardless of the patient being described. Inspection of Figure 1 shows the 20 categories of interpersonal experience arranged in such a way that they define two major dimensions: positive-negative and active-passive. Thus, the positive section of the positive-negative dimension is defined by the 10 categories A through E and K through O on the right side of Figure 1, while the negative section is defined by the 10 categories F through J and P through T on the left. This arrangement is,

of course, arbitrary in the sense that it was based upon the investigator's judgment of the interrelationships among the categories, e.g., the decision to include Category R (distrust, demand, accuse) in the definition of passivity might be challenged. Second, we assumed that these 20 categories of interpersonal experience, regardless of differences among observers or patients to be rated, could be grouped into four clusters to define the two dimensions: active-passive and positive-negative (friendly-hostile). These two dimensions, it will be recalled, are included among the three polarities in terms of which Freud (1957) described the vicissitudes of instinct. In his discussion the two dimensions are used for the description of responses to intrapsychic stimuli (instincts). The studies of Osgood, Suci, and Tannenbaum (1957) and Schaefer (1961) suggest that these are important dimensions of the individual's response to his environment. In short, the work cited above explores the possibility that friendly-hostile and active-passive are the two most important of an invariant system of dimensions which characterize responses to internal and external stimuli.

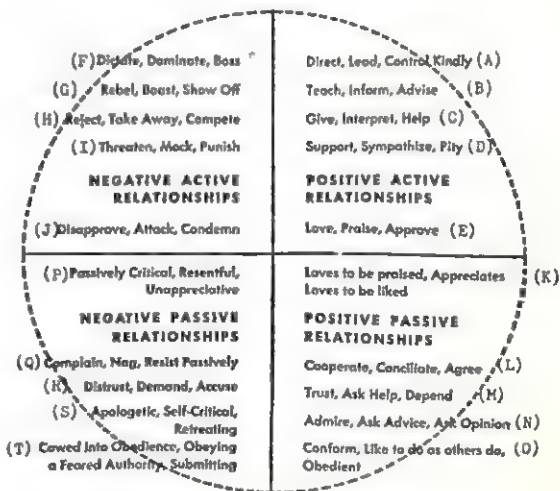


FIG. 1. A system for classifying interpersonal experiences.

<sup>1</sup> This study is part of a major project "Observations and Concepts Used in Clinical Case Description" supported by the National Institute of Mental Health, United States Public Health Service, under Grant M-2224. We are greatly indebted to R. Jarrett, M. B. Smith, and R. Tryon, University of California, for helpful comments and advice on problems of methodology.

The first hypothesis to be tested here is whether this invariant system holds for clinicians with different professional training responding to the same patient and for clinicians with comparable training responding to different patients. If findings support the hypothesis of an invariant system of dimensions, then the second question raised here is whether these dimensions are equally meaningful to all clinical observers and in observations about all patients. In other words, given the fact that the two dimensional system characterizes, in general, the thinking of several groups of clinicians about a patient, what modifications in the use of the system are associated with certain kinds of professional training for the clinician? In what way is the system adapted when professional training is held constant and the patient to be described is varied? In order to spell out the modifications associated with observer or patient differences we would wish to know, first, whether positive-negative and active-passive are two dimensions equally characteristic of clinical thinking regardless of professional training or patient described.

Second, there may still be qualitative differences in the meaning assigned to the positive-negative or the active-passive dimensions. For instance, in the *a priori* system shown in Figure 1, Categories F (dictate, dominate, boss) and A (direct, lead, control kindly) define one pole of the active-passive dimension, while Categories T (cowed into obedience, obeying a feared authority, submitting) and O (conform, like to do as others do, obedient) define the other. If the meaning of this dimension for a given group of clinicians in the context of a given case is the same as the one in the *a priori* system, then Categories F and A should represent the greatest possible contrast to Categories T and O when clinicians describe a patient, regardless of whether this patient is thought to be extremely passive or just average and regardless of the degree of variability in judgments about the patient. However, some observers may consider Category I (threaten, mock, punish) and B (teach, inform, advise) a much better contrast to Categories T and O. In short, differences in the quality of meaning assigned to the two dimensions depend

upon the relationships perceived among the categories which define them. The magnitude of the correlations which measure these relationships depends upon the degree of systematic variation found in the responses of a given sample of observers. Such variation in response may be a function of the particular group of observers studied, of the persons to be rated by the observers, or of a combination of these two factors. In those studies in which responses to a variety of persons are rated by the same observers, the relationships found among the categories are a function of both differences in thinking processes of the observer and differences in the persons rated (Levy & Dugan, 1960; Osgood et al., 1957). In the present study we have held the person to be rated constant in order to evaluate the thinking processes of various groups of observers.

In this study 301 Freudian clinicians were asked to describe the same patients in terms of the categories shown in Figure 1 in order to see if their use of the categories was consistent with the *a priori* system and to determine systematic differences related to professional training and contexts of the clinical material.<sup>2</sup> If the categories are used without special bias by clinicians, one would expect to find that the same dimensions are equally important in their description of the patient and

<sup>2</sup> As described in Chance and Arnold (1960), initially 3,564 clinicians from six major urban areas were contacted. Their names had been obtained from membership directories for the American Psychiatric Association, the Clinical Division of the American Psychological Association, and the Psychiatric Social Work section of the National Association of Social Workers. Willingness to participate was indicated by 1,058 of these clinicians by supplying information about their professional training, experience, and preferred theoretical systems. From this group 542 actually evaluated the clinical material, but only 533 followed instructions closely enough to permit use of their responses in this part of the study. This group of 533 clinicians differed from the sample of 3,564 contacted initially in that psychologists had participated significantly more than members of other disciplines. Of this group 186 psychiatrists, 55 psychologists, and 115 social workers cited Freud as the major influence upon their clinical thinking. In this study only the responses of psychiatrists and social workers were used because the *N* of 55 for psychologists was considered inadequate as stated in Footnote 3.



**Rating Procedure for describing the content of the patient's verbalizations:** The categories listed below should be marked from 1 (least descriptive) to 20 (most descriptive) of the patient's verbalized preoccupation with certain kinds of interpersonal experiences and relationships. First, select out the category which is most characteristic of the patient's verbalizations about his interpersonal experience and assign it the rank of 20. Select the next most characteristic category and assign it the rank of 19. The third most descriptive category should receive the rank of 18. Next, proceed to assign the rank of 1 to the category which least applies to the patient's verbalization about his interpersonal experience. A rank of 2 should be given to the next least descriptive category. Proceed in this fashion until you have used all the ranks from 1 to 20 taking care not to use the same number twice.

We realize that quantification of clinical material raises many problems and is subject to a number of theoretical objections. Nevertheless we hope that for the purpose of this study you will use this procedure to describe your first impressions. Please limit yourself to 15 minutes for rating the transcript in this way.

Twenty Categories of Interpersonal Experience in Random Order\*

- Complain, Nag, Resist passively
- Passively critical, Resentful, Unappreciative
- Reject, Take away, Compete
- Distrust, Demand, Accuse
- Cooperate, Conciliate, Agree
- Teach, Inform, Advise
- Admire, Ask advice, Ask opinion
- Disapprove, Attack, Condemn
- Conform, Like to do as others do, Obedient
- Loves to be praised, Appreciates, Loves to be liked
- Dictate, Dominate, Boss
- Support, Sympathize, Pity
- Rebel, Boast, Show off
- Apologetic, Self-critical, Retreating
- Threaten, Mock, Punish
- Give, Interpret, Help
- Cowed into obedience, Obeying a feared authority, Submitting
- Love, Praise, Approve
- Direct, Lead, Control kindly
- Trust, Ask help, Depend

Your ranking from 1,  
least descriptive, to  
20, most descriptive


Note.—Four hundred randomizations were systematically distributed throughout the sample of 533 clinicians in the original report in order to eliminate the effect of order of presentation.

FIG. 2. Rating sheet used by clinicians.

that each of the categories receives the same empirical definition. Such a design aims at exploring communality in the extent and quality of meaning for clinicians with different professional training. The same procedure in each group of clinicians, replicated for several patients, permits exploration of questions about stability of meaning related to differences among patients.

### METHOD

As reported in an earlier paper (Chance & Arnold, 1960) psychiatrists and social workers from six major urban areas in the United States had been asked to evaluate the transcripts of two treatment sessions: one analytic hour with a middle-aged business man—a sophisticated patient deeply involved in psychoanalysis, and one psychotherapy hour with an ambulatory, schizophrenic girl—promiscuous, alcoholic, and unable to use the hour for more than a report of her daily mishaps. Both patients had been seen by the same therapist for over one year. Mimeographed transcripts of the two treatment

sessions were sent to each participating clinician together with instructions to evaluate the verbalizations of each patient by means of the instrument shown in Figure 2. It will be noted that the categories which the clinicians were asked to rank in order of their descriptiveness for the verbalizations of each patient are the same as those shown in Figure 1, but were presented to the subjects in different random orders. Half of the subjects were instructed to read and evaluate the therapy transcript first; the other half were asked to read and evaluate

TABLE 1.  
BREAKDOWN OF THE SAMPLE

Sample	Number describing	
	Analytic patient	Therapy patient
Analytically trained psychiatrists	54	51
Psychiatrists without analytic training	37	44
Social workers	62	53



the analytic transcript first. This attempt at counterbalancing the effects of order of case presentation was not successful.<sup>3</sup> Therefore, only responses to the first case read were used in this investigation. Table 1 shows the breakdown of the sample of Freudian psychiatrists and social workers who returned the materials in usable condition. One hundred and fifty-three evaluated the analytic transcript, and 148 the therapy transcript. Within each of these major divisions of the sample the responses of three groups were examined separately: psychiatrists with formal analytic training, psychiatrists without such training, and social workers. We were interested, first, in determining whether given the same patient clinicians with varying professional experience would agree in the meaning attributed to the 20 categories. The second problem was to determine whether holding professional training constant clinicians assessing two different cases would attribute the same meaning to the 20 categories of interpersonal relationships. Technical problems in the investigation of these questions can be considered in two steps: first, the choice of a technique for describing the relationships among the ranks obtained from a given group for the 20 categories in terms of underlying dimensions, and, second, the choice of the most logical procedure for comparing the responses of six groups of clinicians.

Factor analysis was the technique used for analyzing the relationships obtained among rankings for the 20 categories by each group of clinicians in terms of the dimensions underlying these rankings. The first step was to factor the covariance matrices generated from the six sets of category rankings. This was done using the centroid method after correcting for the effects of ipsative scores according to Tucker (1956). The factor loadings thus obtained were then standardized by dividing by the standard deviations of the associated variables in order to obtain results equivalent to a factor analysis of correlation coefficients. In line with the first hypothesis concerning the importance of the two dimensions, it was decided that unless the first two factors from a given analysis could account for at least 25% of the variance present, we were not interested in proceeding further. In addition, since the amount of variance left to be extracted was, in general, minimal after removing the first two centroid factors, it was decided to extract only these two and a third to assist in matching them to the criterion.<sup>4</sup>

<sup>3</sup> It was demonstrated that order of case presentation showed marked effects upon the clinicians' judgment about each patient's concern with negative-passive relationships. If a patient was assessed first, he was thought to be more preoccupied with resentment, complaint, and distrust than if he was assessed second.

<sup>4</sup> The factor loadings were stabilized by two separate factorings of the covariance matrices using loadings obtained in the first factoring of categories to estimate the communalities for the second. The differences between the first and second estimates were

In order to clarify the procedures used, the model and the hypothesis are presented in some detail. The main hypothesis is that for all English speaking clinicians there exists an invariant semantic structure associated with the 20 categories presented in Figure 1. That is to say, that within the limits imposed by errors of sampling and measurement, there exists an orthogonal rotation of the factor matrix of one group of clinicians to a position where it will correlate perfectly with the factor matrix of another group of clinicians. Note that this says nothing about the relative amount of variance associated with the various factors. The theoretical correlations are not affected by this. The test of the hypothesis is to rotate one factor matrix to maximum agreement with another and determine whether the correlation between the corresponding factor loadings departs materially from 1.00. A degree of discretion is required here since departure from 1.00 will to some extent be a function of the reliability of the loadings, and loadings on factors that extract smaller proportions of variance may be expected to be less reliable than those on factors that extract larger amounts of variance.

For comparisons of factorial results from any one group with those from another, reference axes for the factor matrix of that group were rotated to a least squares solution<sup>5</sup> of maximum agreement with results for the other group. The comparisons reported here involved the rotations listed in Table 2. In these comparisons the responses of analytically trained psychiatrists to the analytic case were selected as the criterion performance. It was assumed that this particular group of clinicians had the longest and most intensive training and that their responses would be most representative of Freudian clinical thinking. Responses to the analytic transcript were selected as the criterion performance since the case with that patient represented a more classically Freudian instance of therapy than the session with the psychotherapy patient. As shown in Table 2 studies of communality of meaning among observers with varying professional training for the analytic case involved comparison of the responses of psychiatrists without analytic training and of those of social workers with the criterion performance. For studies of communality of meaning in the description of the therapy patient the responses of psychoanalytically trained psychiatrists for that patient were first rotated toward the criterion performance in order to obtain the most "typically" Freudian organization of the data. The responses to the therapy

so slight that the results of the second were accepted, and the procedure was carried no further.

<sup>5</sup> The reference axes were rotated to minimize the squared distances between comparable vectors (variables). The solution requires maximizing the inner products (covariances) of comparable factor vectors. Orthogonal solutions were used. Since we had no hypothesis concerning the relationship between the two factors, orthogonality appeared the simplest and most parsimonious assumption.

TABLE 2

## MATRIX ROTATIONS FOR COMPARING RESULTS OF FACTORIAL STUDIES

Observers	Analytic patient	Therapy patient
Matrix rotations for studies of communality of meaning among observers with varying professional training		
Analytically trained psychiatrists	Matrix 1 Criterion (No rotation)	Matrix 2 Rotated to criterion Matrix 1 (Criterion for other therapy-patient matrices after rotation)
Psychiatrists without analytic training	Matrix 3 Rotated to criterion Matrix 1	Matrix 4 Rotated to criterion Matrix 2
Social workers	Matrix 5 Rotated to criterion Matrix 1	Matrix 6 Rotated to criterion Matrix 2
Matrix rotations for studies of stability of meaning in the context of different patients		
Analytically trained psychiatrists	Matrix 1 Criterion (No rotation)	Matrix 2 Rotated to criterion Matrix 1
Psychiatrists without analytic training	Matrix 3 Rotated to criterion Matrix 1 (Criterion after rotation for therapy-patient matrix of similar group)	Matrix 4 Rotated to criterion Matrix 3
Social workers	Matrix 5 Rotated to criterion Matrix 1 (Criterion after rotation for therapy-patient matrix of similar group)	Matrix 6 Rotated to criterion Matrix 5

transcript of psychiatrists without analytic training and of social workers were then compared with those of the analytically trained psychiatrists.

In order to determine the extent to which this sample of Freudian clinicians attributed the same meaning to the 20 categories in the context of describing the analytic and the therapy transcripts, responses for the analytic patient by psychiatrists without analytic training and by social workers were first rotated toward the criterion performance of analytically trained psychiatrists. Having thus obtained the most "typically" Freudian arrangement of the categories, the responses of each of the three groups of clinicians for the analytic patient were then compared with those of the corresponding groups for the therapy patient.

In all comparisons, regardless of whether they concerned communality of meaning among observers with different professional training or stability of meaning in the description of different patients, the

phenomena of interest can be described in quantitative terms<sup>6</sup> as follows:

1. The importance of the two-dimensional system for the thinking of each group of clinicians can be measured in terms of the average proportion of total variance explained by the positive-negative and the active-passive factors.

2. Whereas the importance of the system as a whole may remain unaffected by variations in the professional training of observers or by variations in clinical material described, the relative importance of the two factors may be affected by such variations. Variations in the ratio of the average propor-

<sup>6</sup> Statistical tests of differences between variance-covariance matrices were not undertaken because the basic data were ranks, thus rendering the matrices only approximations of what might have been obtained with scores with less severe restrictions on them.



tion of the total variance explained by the two factors were, therefore, examined for evidence of the effects of different professional training or different clinical material.

3. Similarity in the meaning of the two dimensions for clinicians with different professional training or for the description of different patients can be measured as follows: The relationship of each category to an underlying factor such as the positive-negative dimension is described by its loading on that factor. Factor loadings obtained for a given matrix for each of the 20 categories on the positive-negative dimension can be correlated with those obtained from the criterion matrix. Product-moment correlations were used.

4. While the above correlations may serve as measures of similarity in the over-all meanings assigned to the dimensions, this method fails to pinpoint the kind of differences in the meaning of the dimensions which might occur among different observers or in the context of different patients. In order to explore such qualitative shifts in meaning, the dimensions were arbitrarily considered to be empirically defined for each group of observers by the five categories at each pole (i.e., those with the highest positive and the highest negative algebraic factor loadings). For example, for the criterion group of psychoanalytically

trained psychiatrists, the five most active and the five most passive categories define the dimension of activity-passivity in the context of a given case. Comparison of this definition of activity-passivity with that produced by a group with similar training in the course of describing the therapy patient reveals shifts in meaning related to shifting clinical context. Comparison of the criterion group's definition of the dimension in the context of the analytic case with definition of the dimension in the same context produced by another group of observers reveals shifts in meaning related to professional training.

## RESULTS AND DISCUSSION

### *Findings<sup>7</sup> Concerning Community of Meaning among Observers with Different Professional Training*

Descriptions of the analytic patient's verbalizations: The results of factoring the responses of the criterion group of analytically trained psychiatrists, of psychiatrists without analytic training, and of social workers are shown in Tables 3, 4, and 5, respectively. It will be noted that in each of these three factor analyses the first two factors extracted were the positive-negative and active-passive dimensions postulated in the system shown in Figure 1. The importance of the two-dimensional system as measured by the average proportion of the total variance explained by the first two factors ranged from .30 for the criterion group to .32 for the group of social workers, failing to differentiate among the three groups.

The relative importance assigned to the two factors did differentiate among clinicians. The criterion group of analytically trained psychiatrists assigned greater importance to the positive-negative than to the active-passive dimension. The social workers assigned equal importance to the two dimensions. In contrast to this psychiatrists without analytic training assigned greater importance to the active-passive dimension than to the positive-negative dimension. The emphasis on the positive-negative relative to the active-passive

TABLE 3

MATRIX OF STANDARDIZED FACTOR LOADINGS FOR RESPONSES BY THE ANALYTICALLY TRAINED PSYCHIATRISTS TO THE ANALYTIC TRANSCRIPT

Category	Factor			$h^2$
	I	II	III	
A (Lead)	.31	.57	.14	.44
B (Teach)	.41	.33	.31	.37
C (Give)	.57	.48	-.22	.60
D (Support)	.44	.42	.17	.40
E (Love)	.46	.21	-.12	.27
F (Boss)	-.23	.49	-.25	.36
G (Rebel)	-.40	.51	.31	.52
H (Reject)	-.39	.30	.33	.35
I (Punish)	-.52	.23	-.20	.36
J (Attack)	-.66	.14	-.43	.64
K (Appreciates)	.28	-.19	-.26	.18
L (Cooperate)	.52	-.36	.27	.47
M (Trust)	.51	-.29	-.19	.38
N (Admire)	.48	-.32	-.25	.40
O (Conform)	.36	-.48	.24	.42
P (Resentful)	-.58	-.26	.43	.59
Q (Complain)	-.42	-.29	.19	.30
R (Accuse)	-.59	-.06	-.25	.41
S (Retreating)	-.13	-.59	-.22	.41
T (Submitting)	-.37	-.54	.16	.45
Sum of squares	4.05	2.93	1.35	
Average proportion of total variance	.18	.12	.06	
Average proportion of common variance	.50	.33	.17	

<sup>7</sup> Statistical data supplementary to this paper have been deposited with the American Documentation Institute. Order Document No. 7092 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$2.00 for microfilm or \$3.75 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.



TABLE 4

MATRIX OF STANDARDIZED FACTOR LOADINGS FOR  
RESPONSES BY THE PSYCHIATRISTS WITHOUT  
ANALYTIC TRAINING TO THE ANALYTIC  
TRANSCRIPT

Category	Rotated factors <sup>a</sup>			h <sup>2</sup>
	I	II	III	
A (Lead)	.02	.39	-.01	.15
B (Teach)	.10	.74	.17	.59
C (Give)	.39	.49	-.29	.48
D (Support)	.25	.06	.11	.08
E (Love)	.23	.16	-.43	.26
F (Boss)	-.49	.59	.03	.59
G (Rebel)	-.16	.37	.14	.18
H (Reject)	-.33	.12	.15	.15
I (Punish)	-.06	.52	-.61	.65
J (Attack)	-.19	.36	-.40	.33
K (Appreciates)	.38	-.33	.46	.46
L (Cooperate)	.60	-.06	-.11	.38
M (Trust)	.37	-.34	.01	.25
N (Admire)	.52	-.24	.15	.35
O (Conform)	-.08	-.48	-.20	.28
P (Resentful)	-.40	-.18	.67	.64
Q (Complain)	-.38	-.31	.46	.45
R (Accuse)	-.72	.12	.13	.55
S (Retreating)	-.14	-.77	.00	.61
T (Submitting)	-.05	-.87	-.18	.79
Sum of squares	2.46	3.88	1.88	
Average proportion of total variance	.12	.19	.09	
Average proportion of common variance	.30	.47	.23	

Transformation matrix

.264282	-.909282	.321498
.961968	.224658	-.155415
.069124	.350344	.934070

<sup>a</sup> Rotated to matrix of analytically trained psychiatrists for analytic transcript.

sive factor can be expressed for the three groups by the following ratios: 1.5, 1.0, and .63. Psychiatrists without analytic training presented an almost exact reversal of emphasis compared with psychiatrists with formal analytic training.

Shifts in the over-all meaning of the two dimensions were affected by differences in the professional training of observers as follows: Correlations between the responses of the criterion group and the psychiatrists without analytic training and social workers were .83 and .90, respectively, for the factor loadings of categories on the positive-negative dimension, and .67 and .92 for those on the active-passive dimension. Although differences

among these correlations cannot be evaluated statistically, the data suggest that the social workers resembled the analytically trained psychiatrists more in their use of these dimensions than psychiatrists without psycho-analytic training.

Empirical definitions of each dimension by each group of observers are illustrated in Figures 3 and 4. Figure 3 shows the five categories considered most positive and the five considered most negative by each group of observers. Inspection of the figure shows that all three groups of clinicians agree in defining this dimension in terms of the polarity: trust, ask help, depend (Category M)-distrust, accuse, demand (Category R).

TABLE 5

MATRIX OF STANDARDIZED FACTOR LOADINGS FOR  
RESPONSES BY THE SOCIAL WORKERS TO  
THE ANALYTIC TRANSCRIPT

Category	Rotated factors <sup>a</sup>			h <sup>2</sup>
	I	II	III	
A (Lead)	.27	.47	.30	.38
B (Teach)	.26	.38	.33	.32
C (Give)	.24	-.02	.19	.09
D (Support)	.43	.31	.21	.33
E (Love)	.54	.23	-.47	.57
F (Boss)	-.24	.79	-.24	.74
G (Rebel)	-.33	.56	.14	.44
H (Reject)	-.52	.26	-.21	.38
I (Punish)	-.49	.40	-.32	.50
J (Attack)	-.81	.07	-.02	.66
K (Appreciates)	.49	-.12	-.09	.26
L (Cooperate)	.21	-.41	.34	.33
M (Trust)	.28	-.41	-.23	.30
N (Admire)	.27	-.46	-.16	.31
O (Conform)	.48	-.40	.03	.39
P (Resentful)	-.27	-.36	.55	.51
Q (Complain)	-.21	-.40	.41	.37
R (Accuse)	-.64	-.05	-.38	.56
S (Retreating)	.02	-.64	-.03	.41
T (Submitting)	.01	-.36	-.16	.16
Sum of squares	3.20	3.24	1.57	
Average proportion of total variance	.15	.16	.08	
Average proportion of common variance	.40	.40	.20	

Transformation matrix

.647449	-.750007	.135273
.748315	.659254	.073546
-.144339	.053608	.988075

<sup>a</sup> Rotated to matrix of analytically trained psychiatrists for analytic transcript.

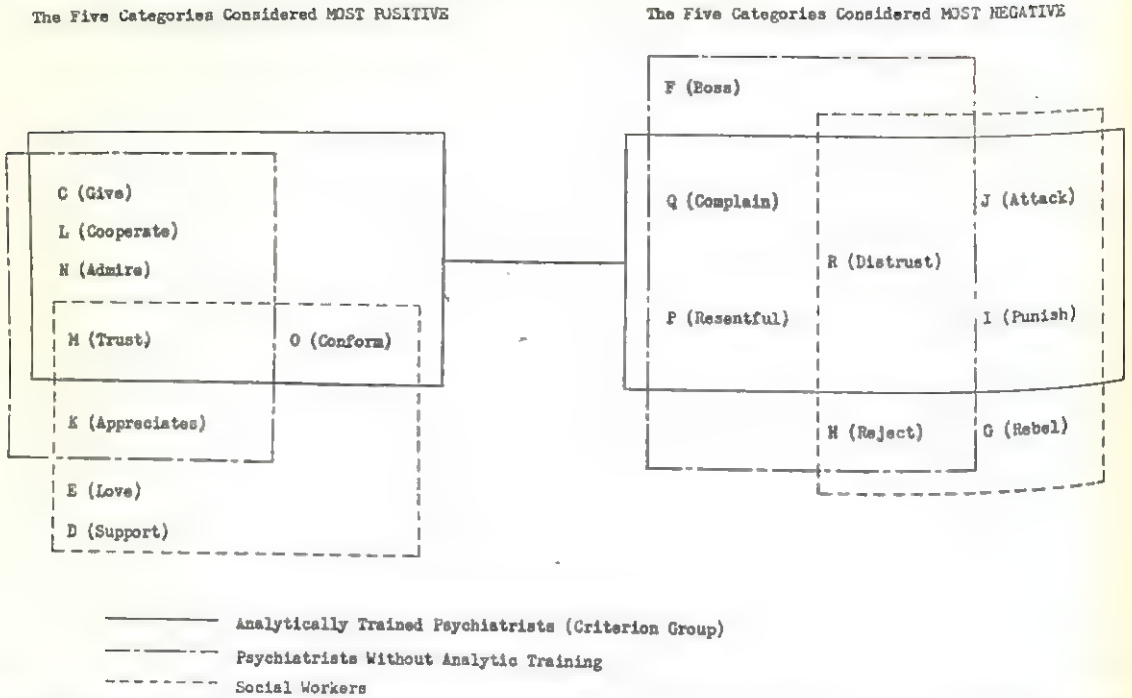


FIG. 3. Comparisons of empirical definitions of the positive-negative dimension in the context of the analytic patient.

Figure 4 shows empirical definitions for the active-passive dimension. Again it will be noted that of the 10 categories which consti-

tute the criterion group's definition of this dimension all but 1 were also used by either one or both of the other two groups of cli-

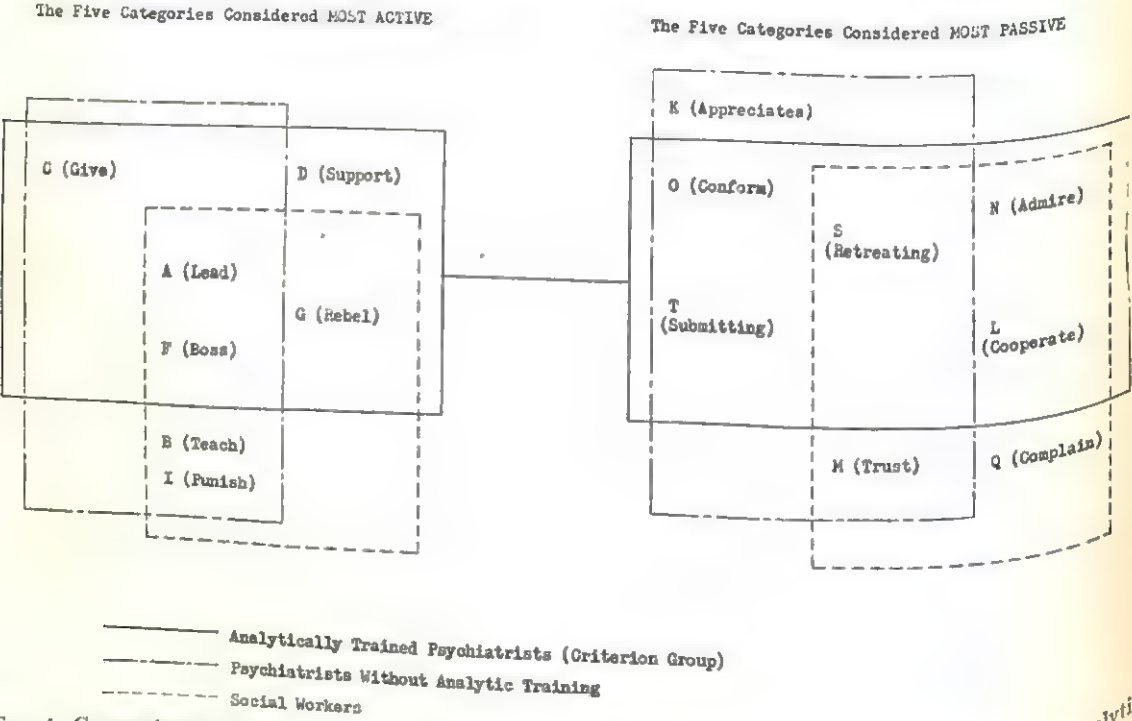


FIG. 4. Comparisons of empirical definitions of the active-passive dimension in the context of the analytic patient.

TABLE 6

MATRIX OF STANDARDIZED FACTOR LOADINGS FOR  
RESPONSES BY THE ANALYTICALLY TRAINED  
PSYCHIATRISTS TO THE THERAPY  
TRANSCRIPT

Category	Rotated factors <sup>a</sup>			h <sup>2</sup>
	I	II	III	
A (Lead)	.56	.31	.21	.45
B (Teach)	.28	.54	.28	.45
C (Give)	.63	.40	.11	.57
D (Support)	.49	.30	-.21	.37
E (Love)	.55	.12	-.31	.41
F (Boss)	-.43	.19	.17	.25
G (Rebel)	-.42	.54	-.22	.52
H (Reject)	-.57	-.02	.30	.42
I (Punish)	-.49	.27	-.30	.40
J (Attack)	-.63	-.02	.05	.40
K (Appreciates)	.32	-.11	-.48	.34
L (Cooperate)	.36	-.28	.13	.22
M (Trust)	.47	-.26	.25	.35
N (Admire)	.35	-.29	.15	.23
O (Conform)	.32	-.59	-.11	.46
P (Resentful)	-.49	-.08	.19	.28
Q (Complain)	-.67	-.06	.16	.48
R (Accuse)	-.72	-.03	.19	.56
S (Retreating)	-.02	-.62	-.18	.42
T (Submitting)	.11	-.30	-.23	.16
Sum of squares	4.55	2.13	1.06	
Average proportion of total variance	.23	.11	.05	
Average proportion of common variance	.59	.27	.14	

## Transformation matrix

.976077	-.190091	-.105533
.163094	.961131	-.222772
.143778	.200230	.969142

<sup>a</sup> Rotated to matrix of analytically trained psychiatrists for analytic transcript.

nicians. The use of Category D (support, sympathize, pity) as part of their definition of activity distinguished analytically trained psychiatrists from the other two groups. Both social workers and psychiatrists without analytic training differed from the criterion group in using Categories B (teach, inform, advise) and I (threaten, mock, punish) in their definition of activity. Both also differed from the criterion group in that they included Category M (trust, ask help, depend) in their definition of passivity. In summary, empirical definitions of the active-passive dimension appeared in the context of the analytic case to differentiate more sharply between analytically trained and other clinicians than the

definitions of the positive-negative dimension.

Descriptions of the therapy patient's verbalizations: The results of factoring responses of the criterion group of analytically trained psychiatrists, of psychiatrists without analytic training, and of social workers are shown in Tables 6, 7, and 8. Findings for responses to the therapy patient were similar to those for the analytic patient in that the first two factors extracted in each of three factor analyses were the positive-negative and active-passive dimensions postulated in the a priori system. The importance of the two-dimensional system as measured by the average proportion of total variance explained by the first two factors represented a wider range: .34

TABLE 7

MATRIX OF STANDARDIZED FACTOR LOADINGS FOR  
RESPONSES BY THE PSYCHIATRISTS WITHOUT  
ANALYTIC TRAINING TO THE  
THERAPY TRANSCRIPT

Category	Rotated factors <sup>a</sup>			h <sup>2</sup>
	I	II	III	
A (Lead)	.15	.47	.14	.26
B (Teach)	.54	.47	.04	.51
C (Give)	.72	-.04	.22	.57
D (Support)	.59	-.29	.33	.54
E (Love)	.84	.05	-.03	.71
F (Boss)	-.64	.40	.49	.81
G (Rebel)	-.49	.47	-.07	.47
H (Reject)	-.75	.28	-.05	.64
I (Punish)	-.64	.10	-.26	.49
J (Attack)	-.81	-.03	.15	.68
K (Appreciates)	.37	-.20	-.62	.56
L (Cooperate)	.41	-.25	-.23	.28
M (Trust)	.78	-.05	-.06	.61
N (Admire)	.67	.22	-.08	.50
O (Conform)	.39	-.28	-.30	.32
P (Resentful)	-.69	-.12	.33	.60
Q (Complain)	-.53	-.30	.38	.52
R (Accuse)	-.73	.26	-.18	.63
S (Retreating)	.19	-.42	.02	.21
T (Submitting)	.06	-.32	-.03	.11
Sum of squares	7.01	1.68	1.34	
Average proportion of total variance	.35	.08	.07	
Average proportion of common variance	.70	.17	.13	

## Transformation matrix

.938945	-.298356	-.171365
.342764	.767823	.541258
-.029910	-.566949	.823210

<sup>a</sup> Rotated to rotated matrix of analytically trained psychiatrists for therapy transcript.



TABLE 8

MATRIX OF STANDARDIZED FACTOR LOADINGS FOR RESPONSES BY THE SOCIAL WORKERS TO THE THERAPY TRANSCRIPT

Category	Rotated factors <sup>a</sup>			h <sup>2</sup>
	I	II	III	
A (Lead)	.41	.13	.06	.19
B (Teach)	.20	.12	-.30	.14
C (Give)	.67	.05	.15	.47
D (Support)	.64	-.17	.25	.50
E (Love)	.64	.00	-.27	.48
F (Boss)	-.73	.04	-.02	.53
G (Rebel)	-.52	.18	.15	.33
H (Reject)	-.61	.18	-.11	.42
I (Punish)	-.64	.35	-.25	.59
J (Attack)	-.67	.35	.12	.59
K (Appreciates)	.43	.35	-.23	.36
L (Cooperate)	.41	-.40	-.10	.34
M (Trust)	.71	.21	.11	.56
N (Admire)	.64	.01	.24	.47
O (Conform)	.31	-.44	-.39	.44
P (Resentful)	-.55	-.10	-.09	.32
Q (Complain)	-.61	-.27	.30	.54
R (Accuse)	-.63	.22	.28	.52
S (Retreating)	.28	-.40	.29	.32
T (Submitting)	-.10	-.20	-.17	.08
Sum of squares	6.03	1.23	.94	
Average proportion of total variance	.30	.06	.05	
Average proportion of common variance	.74	.15	.11	

Transformation matrix		
.990466	-.128751	-.048996
.130223	.991086	.028128
.044938	-.034277	.998403

<sup>a</sup> Rotated to rotated matrix of analytically trained psychiatrists for therapy transcript.

for the criterion group, .36 for social workers, and .43 for psychiatrists without analytic training. In their use of the two-dimensional system social workers resembled analytically trained psychiatrists more than did psychiatrists without analytic training.

In the description of the therapy patient by each group of clinicians the emphasis on the positive-negative dimension was greater than that on the active-passive dimension. However, there was marked contrast between analytically trained psychiatrists and the other two groups in the relative emphasis placed upon this dimension. The ratio for the criterion group was 2.1, whereas it was 4.4 and 5.0, respectively, for psychiatrists without analytic training and for social workers.

The over-all meaning of the positive-negative dimension was not affected by differences in the professional training of clinicians. Correlations of .94 and .96 were obtained between the responses of the criterion group and psychiatrists without analytic training on the one hand and social workers on the other. In contrast to this, the over-all meaning of the active-passive dimension was considerably affected by differences in the professional training of observers. The correlation between the responses of psychiatrists with and without analytic training was .65; that for the responses of social workers with the criterion group was .53.

An examination of the empirical definitions of each dimension by the three groups of clinicians reveals the areas of agreement as well as differences in the quality of meaning attributed to the two dimensions in the context of describing the therapy patient's verbalizations. Figure 5 shows that of the 10 categories used by the criterion group in their definition of the positive-negative dimension all but 1 were used by either or both of the other two groups. As might be expected in view of the high correlations among factor loadings all three groups agreed in defining the positive pole of the dimension in terms of 4 of the 5 categories used for definition of that pole. Agreement concerning the definition of the negative pole was less striking. Empirical definitions of the active-passive dimension are shown in Figure 6. Of the 10 categories used by the criterion group in defining this dimension 8 were also used by either or both of the other two groups. Clinicians with different professional training showed marked differences in defining this dimension. Social workers used an entirely different set of 5 categories from those employed by the other two groups in their definition of activity, but in defining passivity they agreed with the criterion group in the choice of 4 out of 5 categories. While psychiatrists without analytic training agreed with their analytically trained colleagues in the choice of 3 of the 5 categories defining each pole, there was one instance of striking disagreement between the two groups. While analytically trained psychiatrists included Category D (support, sympathize,

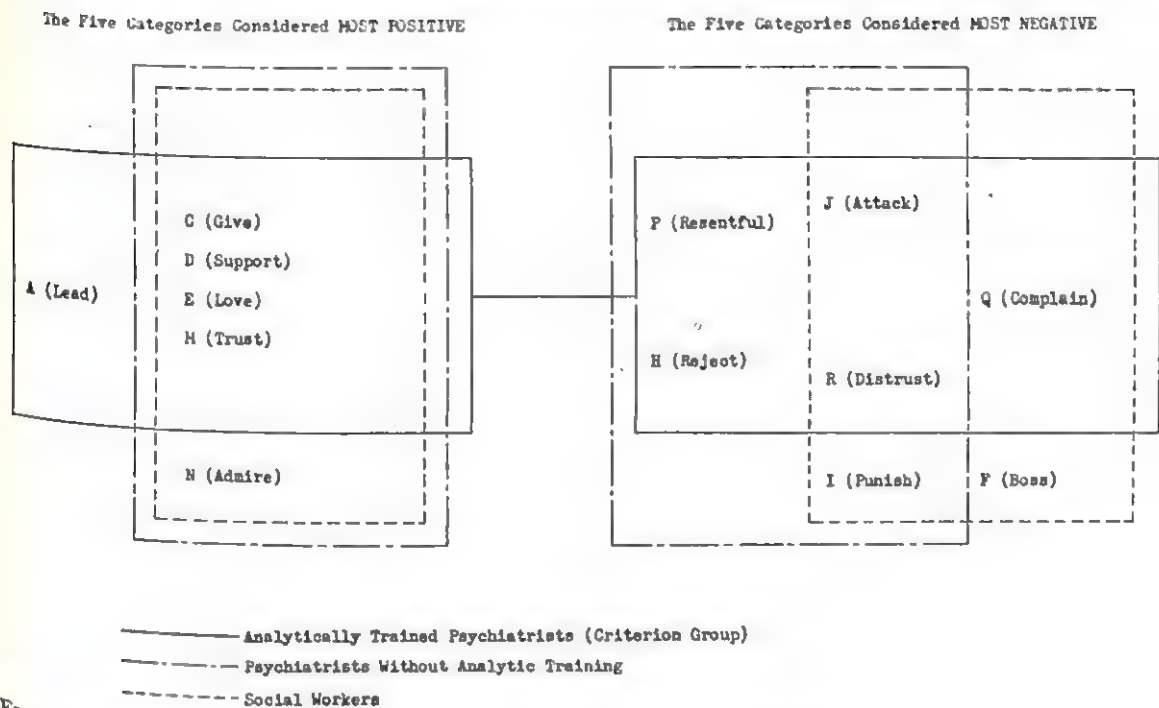


FIG. 5. Comparisons of empirical definitions of the positive-negative dimension in the context of the therapy patient.

pity) in their definition of activity, psychiatrists without analytic training included the same category in their definition of passivity.

Table 9 summarizes quantitative findings concerning communality of meaning among clinicians with different professional training.

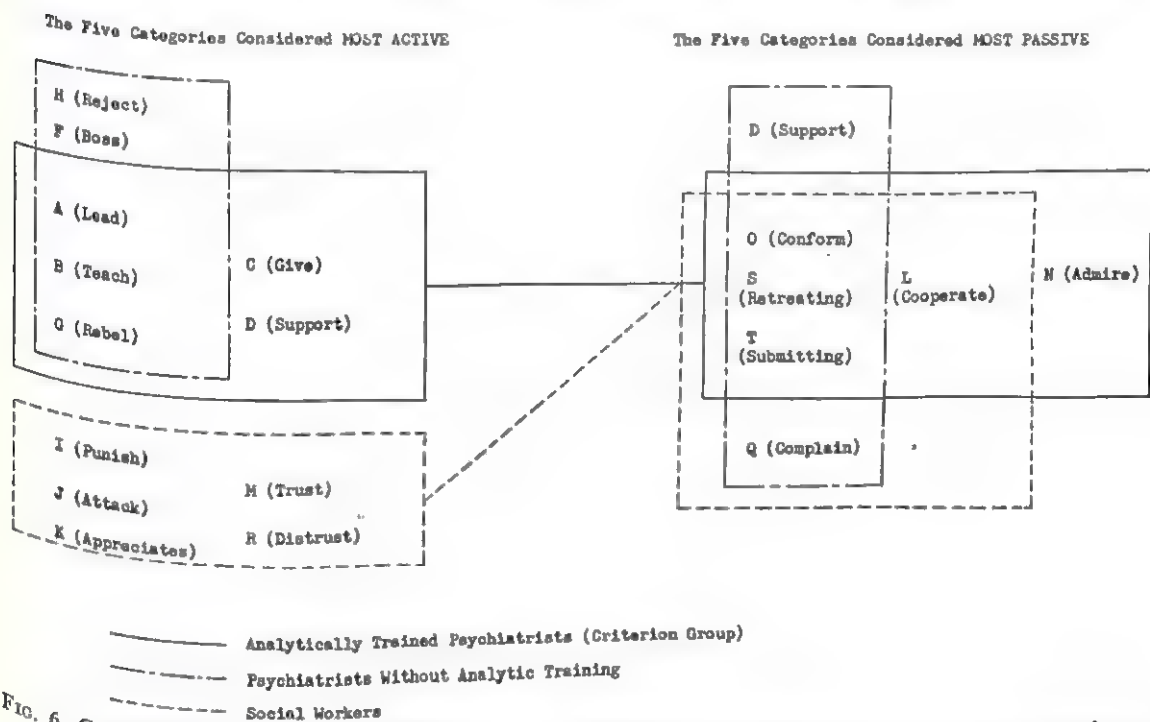


FIG. 6. Comparisons of empirical definitions of the active-passive dimension in the context of the therapy patient.

TABLE 9  
SUMMARY OF FACTORIAL STUDIES COMPARING RESPONSES OF CLINICIANS WITH DIFFERENT  
PROFESSIONAL TRAINING

	Average proportion of total variance explained by the first two factors							
	Analytic patient				Therapy patient			
	Positive-Negative	Active-Passive	Both factors	Ratio <sup>a</sup>	Positive-Negative	Active-Passive	Both factors	Ratio <sup>a</sup>
Criterion group (Analytically trained psychiatrists)	.18	.12	.30	1.5	.23	.11	.34	2.1
Psychiatrists without analytic training	.12	.19	.31	.63	.35	.08	.43	4.4
Social workers	.16	.16	.32	1.0	.30	.06	.36	5.0
	Communality of meaning for each factor measured by correlation between factor loadings							
	Analytic patient				Therapy patient			
	Positive-Negative		Active-Passive		Positive-Negative		Active-Passive	
Criterion group versus Psychiatrists without analytic training	.83		.67		.94		.65	
Criterion group versus Social workers	.90		.92		.96		.53	

<sup>a</sup> Ratio: Positive-Negative/Active-Passive.

The table shows, first, that the postulated two factor system was indeed meaningful to each group of clinicians.

Second, the table shows a tendency in four of the six groups to emphasize the positive-negative rather than the active-passive dimension. This trend is in line with Osgood's (Osgood et al., 1957) findings concerning the relative importance of three dimensions in descriptions of a large variety of concepts: "evaluation," "potency," and "activity." Regardless of the scales used, the concepts described, or the subjects studied, Osgood finds that "differentiation among concepts in terms of their evaluation is about twice as fine as differentiation in terms of their potency or activity" (p. 73). This same trend can be observed in 16 of the 19 separate factor analyses for descriptions of different concepts reported by Osgood et al. (pp. 181-185).

Third, contrary to Osgood's (Osgood et al., 1957, p. 223) findings concerning the effects of subject selection, the table shows that clinicians differed in the relative emphasis placed

upon the two dimensions. These differences were a function of both professional training and the case described. In the context of describing the analytic patient social workers performed more as analytically trained psychiatrists, while in the context of describing the therapy patient they performed more as psychiatrists without analytic training. This apparent tendency to a shift in emphasis may be related to the fact that social workers have opportunities to work with and to learn to adapt to the methods of case description expounded by psychiatrists with and without analytic training. The fact that they responded more as analysts in the analytic case and more as psychiatrists without analytic training in the therapy case may be the result of this capacity to adapt to their professional environment.

Fourth, examination of the table shows that the active-passive dimension, which accounted for a smaller proportion of the variation in responses in four of the six groups of clinicians, is also the dimension which tends



to be used differently by analytically trained and untrained clinicians. Agreement upon the over-all meaning of the dimension ranged from .53 to .92 with three of the four correlations obtained at .67 or below.

Empirical definitions of each dimension in terms of the five categories assigned to each of the poles were examined in order to determine whether the six groups differed in the quality of meaning attributed to the dimensions. It was not surprising that for this sample of Freudian clinicians the groups of analytically trained psychiatrists provided the most generally agreed on definitions. However, in their definition of activity for each case analytically trained psychiatrists consistently differed from the other two groups by including Category D (support, sympathize, pity). Psychiatrists without analytic training consistently used this category as more indicative of passivity and in describing the therapy case even included it among the five most passive attributes which constitute their empirical definition of passivity. A speculative explanation of this difference may be found in examining the definitions of activity which distinguished analytically trained and untrained clinicians. In the context of the analytic case both psychiatrists without analytic training and social workers differed from the criterion group in their inclusion of Categories B (teach, inform, advise) and I (threaten, mock, punish) in definitions of activity. It is possible that Category D (support, sympathize, pity) was contrasted with Categories B and I in the context of the analytic material because clinicians without analytic training may see the psychoanalyst primarily in a passive role.

Definitions of the active-passive dimension in the context of the therapy material presented a more complex picture than in the context of the analytic material. The three groups of clinicians appeared to differ more in their definitions of the active than of the passive pole for this dimension. Whereas the definitions of activity obtained from psychiatrists with and without analytic training contained some of the same categories, the social workers provided an entirely different definition from that produced by the other two groups. Since the therapy patient was a

woman and since the two groups of psychiatrists differed from the social workers not only in professional training but also in terms of a marked predominance of males, one questions whether differences in the definition of activity in the context of material from a woman patient may not be responsible for the differences in the social workers' definition of activity. Definitions of passivity in the context of the therapy material show more agreement among clinicians. However, exclusion of Category Q (complain, nag, resist passively) in the definition of passivity by the criterion group and the inclusion of this category in the definitions obtained from clinicians without analytic training suggests that the latter tend to think of passivity in terms of the manifestly passive meaning of the category, while analytically trained psychiatrists may be more aware of the potential for controlling others through complaints and passive resistance.

In summary, it was possible to demonstrate that using the categories shown in Figure 1 clinical case description, like descriptions of maternal and child behavior (Schaefer, 1961) or descriptions of concepts in general (Osgood et al., 1957), is characterized by two dimensions labeled here positive-negative and active-passive. Communality of meaning for Freudian clinicians with different professional training was greater for the positive-negative than for the active-passive dimension. Differences in meaning were related partly to differences in professional training and partly to the nature of the clinical material. In order to explore the effects of shifts in clinical context upon the stability of meaning of the two dimensions, the responses of the six groups were further analyzed as reported below.

#### *Findings Concerning Stability of Meaning in the Context of the Analytic and the Therapy Patients' Verbalizations*

As outlined in Table 2 studies of stability of meaning in the evaluation of the two patients by the six groups of clinicians involved the following procedures: As a first step, the factor matrices for responses to the analytic data obtained from psychiatrists without analytic training and from social workers were

TABLE 10

MATRIX OF STANDARDIZED FACTOR LOADINGS FOR  
RESPONSES BY THE PSYCHIATRISTS WITHOUT  
ANALYTIC TRAINING TO THE THERAPY  
TRANSCRIPT

Category	Rotated factors <sup>a</sup>			h <sup>2</sup>
	I	II	III	
A (Lead)	.20	.48	.01	.27
B (Teach)	.57	.41	-.09	.50
C (Give)	.72	-.04	.21	.56
D (Support)	.58	-.25	.39	.55
E (Love)	.84	-.03	-.06	.71
F (Boss)	-.59	.56	.39	.81
G (Rebel)	-.46	.48	-.18	.47
H (Reject)	-.73	.32	-.11	.63
I (Punish)	-.64	.09	-.26	.49
J (Attack)	-.81	.08	.16	.69
K (Appreciates)	.33	-.38	-.56	.57
L (Cooperate)	.38	-.33	-.16	.28
M (Trust)	.78	-.13	-.05	.63
N (Admire)	.69	.14	-.14	.52
O (Conform)	.35	-.38	-.23	.32
P (Resentful)	-.69	.03	.36	.61
Q (Complain)	-.54	-.15	.45	.52
R (Accuse)	-.71	.27	-.23	.63
S (Retreating)	.15	-.42	.12	.21
T (Submitting)	.04	-.32	.06	.11
Sum of squares	6.84	1.93	1.32	
Average proportion of total variance	.34	.10	.07	
Average proportion of common variance	.67	.20	.14	

## Transformation matrix

.904651	-.412585	-.106680
.423888	.845414	.324951
-.043882	-.339189	.939695

<sup>a</sup> Rotated to rotated matrix of psychiatrists without analytic training for analytic transcript.

equated as nearly as possible with the factor matrix for responses from the criterion group of analytically trained psychiatrists.<sup>8</sup> These three factor matrices were then used as criterion matrices in order to evaluate shifts in meaning of the two dimensions associated with the assessment of the therapy patient. Factor matrices for responses to the therapy data were, therefore, rotated to these three matrices as shown in Tables 6, 10, and 11.

Findings concerning stability of meaning attributed to the two dimensions in the context of the analytic and the therapy data are summarized in Table 12. Reference to this

<sup>8</sup> Results of these rotations are shown in Tables 3, 4, and 5.

table shows, first, that regardless of differences in professional training, the two-dimensional system accounted on the average for a larger proportion of the total variation in response in the context of the therapy data. This suggests the hypothesis that the system of classifying interpersonal experiences may be more descriptive of therapy data than of analytic data.

Second, the table shows that regardless of differences in professional training, differences in clinical context involved shifts in the relative emphasis placed upon the two dimensions. In the context of the therapy data the positive-negative dimension explained more and the active-passive dimension less of the total average variation in response

TABLE 11

MATRIX OF STANDARDIZED FACTOR LOADINGS FOR  
RESPONSES BY THE SOCIAL WORKERS TO THE  
THERAPY TRANSCRIPT

Category	Rotated factors <sup>a</sup>			h <sup>2</sup>
	I	II	III	
A (Lead)	.42	.09	.09	.19
B (Teach)	.25	.07	-.28	.15
C (Give)	.65	.00	.22	.47
D (Support)	.58	-.21	.33	.49
E (Love)	.67	-.09	-.19	.54
F (Boss)	-.72	.11	-.11	.32
G (Rebel)	-.51	.24	.07	.41
H (Reject)	-.57	.23	-.19	.60
I (Punish)	-.57	.39	-.35	.57
J (Attack)	-.63	.42	.01	.37
K (Appreciates)	.49	.29	-.21	.34
L (Cooperate)	.37	-.45	-.02	.56
M (Trust)	.71	.15	.17	.47
N (Admire)	.61	-.03	.31	.45
O (Conform)	.30	-.51	.31	.32
P (Resentful)	-.55	-.05	-.14	.53
Q (Complain)	-.66	-.18	.25	.51
R (Accuse)	-.62	.30	.18	.31
S (Retreating)	.20	-.39	.35	.08
T (Submitting)	-.10	-.21	-.16	
Sum of squares	5.79	1.41	.98	
Average proportion of total variance	.29	.07	.05	
Average proportion of common variance	.71	.17	.12	

## Transformation matrix

.970441	-.228096	.078840
.232063	.971643	-.045364
-.066257	.062319	.995854

<sup>a</sup> Rotated to rotated matrix of social workers for analytic transcript.

TABLE 12

SUMMARY OF FACTORIAL STUDIES COMPARING RESPONSES OF CLINICIANS IN THE DESCRIPTION OF TWO PATIENTS

Average proportion of total variance explained by the first two factors								
	Analytic patient				Therapy patient			
	Positive-Negative	Active-Passive	Both factors	Ratio <sup>a</sup>	Positive-Negative	Active-Passive	Both factors	Ratio <sup>a</sup>
Criterion group (Analytically trained psychiatrists)	.18	.12	.30	1.5	.23	.11	.34	2.1
Psychiatrists without analytic training	.12	.19	.31	.6	.34	.10	.44	3.4
Social workers	.16	.16	.32	1.0	.29	.07	.36	4.1
	Stability of meaning for each factor measured by correlation between factor loadings							
	Positive-Negative				Active-Passive			
Analytically trained psychiatrists	.94				.91			
Psychiatrists without analytic training	.83				.76			
Social workers	.90				.54			

<sup>a</sup> Ratio: Positive-Negative/Active-Passive.

than in the context of the analytic data. These shifts in emphasis are most plausibly explained by reference to the fact that the analytic patient was a man and the therapy patient a woman. In the assessment of female patients clinicians may attribute greater importance to the positive-negative dimension and less importance to the active-passive dimension.

Third, the table shows that regardless of differences in professional training the overall meaning attributed to the positive-negative dimension was less dependent upon differences in clinical context than that attributed to the active-passive dimension.

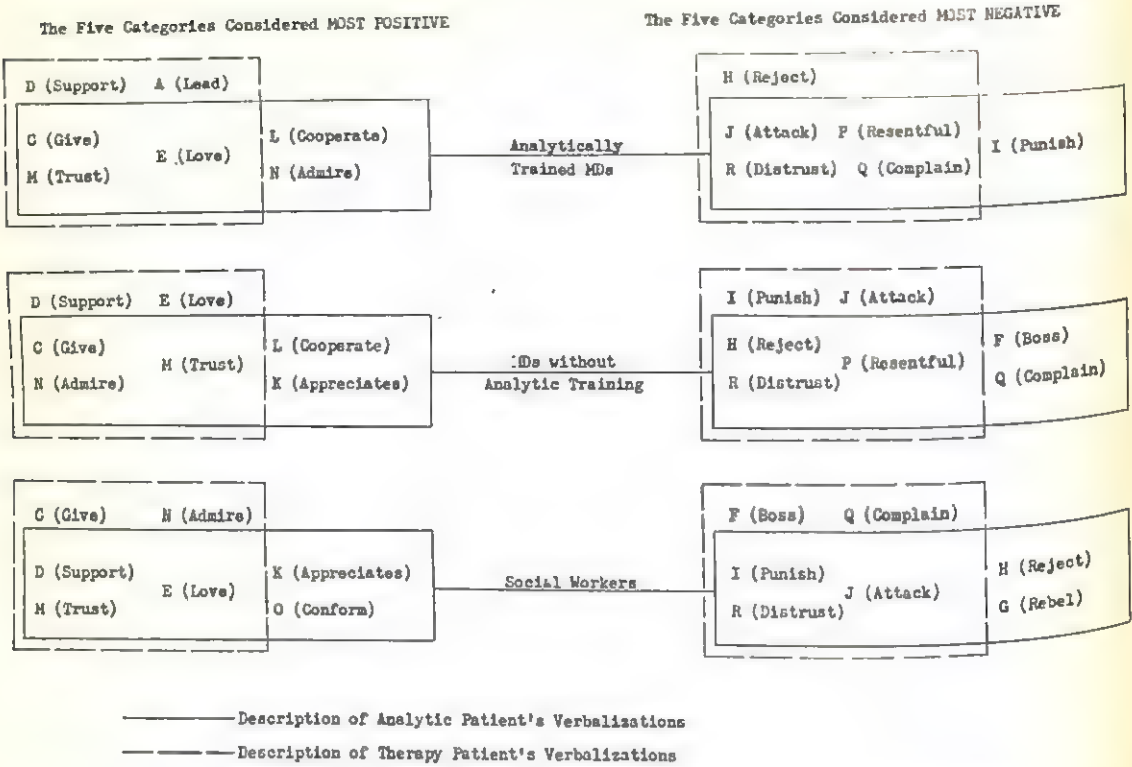
Fourth, the table suggests that differences in professional training may be associated with differences in the stability of meaning attributed to the two dimensions. Whereas correlations between factor loadings for responses to the analytic and the therapy data were .94 and .91 for the criterion group of analytically trained psychiatrists, corresponding correlations obtained for data from psychiatrists without analytic training and from social workers were lower.

The extent to which the quality of mean-

ing attributed to each dimension was affected by differences in clinical context can be evaluated by examining the empirical definitions obtained from each of the six groups of clinicians. The five categories assigned to each pole of the positive-negative dimension by the six groups are shown in Figure 7. Reference to the figure shows, first, that regardless of differences in professional training and differences in clinical context Category M (trust, ask help, depend) defined the positive pole and Category R (distrust, demand, accuse) defined the negative pole of the dimension. Thus, the meaning of the positive-negative dimension was defined in terms of the patient's readiness to relate to people, presumably particularly to therapists.

Empirical definitions of the active-passive dimension are shown in Figure 8. The figure shows that only the passive pole of this dimension could be defined regardless of differences in professional training or clinical context and that passivity was designated by a single category, S (retreating, apologetic, self-critical). Groups of clinicians differed in the extent to which stability of the quality of meaning attributed to this dimension was





affected by shifts in clinical context. Of the 10 categories which, in the responses of each group, were assigned the highest active or passive factor loadings, 8 identical categories were used by the criterion group of analytically trained psychiatrists in the description of each case, 6 by psychiatrists without training, and 4 by social workers. The relative immunity of the definition obtained from analytically trained psychiatrists to shifts in clinical context involving sex differences suggests that training in analytic theory (particularly training in the theory of psychosexual development) may contribute to a more generally applicable conceptualization of the active-passive dimension.

In reviewing these findings it should be recalled that whereas the first hypothesis concerning invariant characteristics of clinical thinking postulates that clinicians think alike in that they share a common frame of reference, the second hypothesis postulates that modifications in this common frame of reference may be associated with differences in professional training or differences in the patient described. The first hypothesis was supported by our findings wherever there was high correlation between factor loadings for the two-factor systems produced by each of the groups. When such high correspondence is found in the meaning of the two dimensions, we can have considerable faith in quantitative differences between the groups in terms of the over-all importance assigned to the two-factor system and the relative importance of the two factors. In other words, the second hypothesis concerning modifications of the two-factor system associated with professional training or patient described can be explored in quantitative terms, and qualitative differences should not receive too much attention. Where the first hypothesis concerning the invariant structure of clinical thinking found only moderate support in that correlations between factor loadings for the two factors were relatively low, quantitative comparisons of groups in terms of the over-all importance of the two-factor system or the relative importance of the two factors in each group rest on insecure ground. In this case qualitative modifications of the system, e.g., in the social workers' definition of ac-

tivity for the female patient, should receive greater attention.

### SUMMARY

The purpose of this investigation was to examine the influence of professional training and clinical material upon the meaning of concepts used in case description. Three hundred and one Freudian clinicians were asked to describe the patients' verbalizations in the transcripts of one psychoanalytic and one psychotherapy session by means of 20 randomly presented categories of interpersonal experience. These categories had, in earlier research, been classified in terms of two dimensions derived from analytic theory: positive-negative and active-passive. Responses obtained from these subjects were treated by six separate factor analyses, one each for responses to the analytic and the therapy transcripts by psychiatrists with analytic training, by psychiatrists without analytic training, and by social workers.

Results included the following findings:

1. The two-dimensional system developed a priori was found to characterize clinical case description in that the first two factors underlying the responses of each group represented the postulated dimensions.
2. The two-dimensional system appeared to be more descriptive of the psychotherapy patient than of the analytic patient.
3. In the description of the psychotherapy patient the positive-negative dimension received more and the active-passive dimension received less emphasis than in the description of the analytic patient.
4. Clinicians differed according to their professional training in the relative importance attributed to the two dimensions. In the context of the analytic data, social workers responded more as psychoanalysts, while in the context of the therapy data they responded more as psychiatrists without analytic training.
5. The clinicians differed in the extent to which the meaning attributed to the two dimensions remained stable regardless of the patient described. The responses of psychoanalysts showed the most stability of the three groups.

6. The over-all meaning of the positive-negative dimension was found to be more independent of shifts in clinical context or of differences in professional training of subjects than that of the active-passive dimension.

7. Whereas the positive-negative dimension, regardless of differences in professional training or clinical context, was defined in terms of the polarity trust-distrust, definitions of the active-passive dimension varied with professional training and clinical context, particularly with respect to the active pole of the dimension.

#### REFERENCES

- CHANCE, ERIKA. *Families in treatment*. New York: Basic Books, 1959.
- CHANCE, ERIKA, & ARNOLD, J. The effect of professional training, experience, preference for a theo-

retical system upon clinical case description. *Hum. Relat.*, 1960, 13, 195-213.

FREUD, S. *The standard edition of the complete psychological works of Sigmund Freud*. Vol. XIV. (Originally published 1914-1916) London: Hogarth, 1957. P. 127.

LEVY, L. H., & DUGAN, R. D. A constant error approach to the study of dimensions of social perception. *J. abnorm. soc. Psychol.*, 1960, 61, 21-24.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univ. Illinois Press, 1957.

SCHAEFER, E. Converging conceptual models for maternal behavior and for child behavior. In J. Glidewell (Ed.), *Parental attitudes and child behavior*. Springfield, Ill.: Charles C Thomas, 1961, in press.

TUCKER, L. Factor analysis of double centered score matrices. Princeton: Educational Testing Service, 1956.

(Received April 3, 1961)



## EGO INVOLVEMENT:

### A CRITICAL EXAMINATION OF SOME METHODOLOGICAL ISSUES

EVA DREIKURS FERGUSON<sup>1</sup>

*Craig House for Children, Pittsburgh, Pennsylvania*

As psychologists have increasingly turned to the laboratory investigation of human motivation, a host of empirical and conceptual problems has arisen. At the present time, various operational methods for obtaining different degrees of motivation exist, and a variety of types of motivations are being investigated, e.g., anxiety, need achievement, failure-stress, ego involvement. One method of obtaining high and low degrees of a given motivation is to use subjects who differ in their degree of this motivation prior to entry into the laboratory. Another method is to induce different degrees of motivation in the laboratory, in subjects whose existent level of the motivation under study was not ascertained. Although a combination of these methods may be utilized, the use of the latter method has been commonly employed to elicit ego involvement. By appropriate instructions or other means of orienting subjects to the experimental situation, different degrees of an ego involvement motivation are assumed to be established. However, unless a direct measure is made of the extent to which instructions in fact evoke ego involvement motivation, data relating differences in instructions to differences in behavior on performance or learning tasks cannot be readily interpreted. A lack of performance effects from ego involving instructions may be due to the fact that ego involvement motivation was not altered by the instructions, or it may be due to

the fact that although instructions produced a change of motivation, this latter change did not lead to the predicted performance changes. Additional problems become apparent from a review of the literature on ego involvement. Stimuli used to evoke ego involvement ( $EI_s$ ) are not uniform. Tasks presented to subjects and performance measures used to test the effects of  $EI_s$  vary greatly. Although some attention has been paid to exploring the stimulus dimension of  $EI_s$  (e.g., Shedd & Angelino, 1952), relatively little work has been done in this area. And as yet there appears to be no direct response measure for ego involvement ( $EI_R$ ), which would eventually allow assessing the degree to which various  $EI_s$  in fact evoke the assumed ego involvement motivation.

At least in part as a result of these factors, data on the effects of ego involvement on behavior are inconclusive. Some investigators have shown ego involvement to affect learning (Alper, 1946; Heyer & O'Kelly, 1949) while others have shown it not to do so (Kausler, 1951; Russell, 1952). Where performance has been affected by ego involvement it is sometimes improved (e.g., Kausler, 1951) and sometimes deteriorated (Castaneda & Palermo, 1955; Romanow, 1958; Shedd & Angelino, 1952). Other studies show ego involvement to have no effect (e.g., Irwin & Mintzer, 1942), or to have only a minimal effect (Alper, 1957; Holt, 1945; Vogel, Raymond, & Lazarus, 1959).

A close examination of certain important issues may lead to an ultimate unraveling of the inconclusive nature of present findings in the ego involvement literature. Whether ego involving instructions alter motivation has been answered in the affirmative by two studies, even though neither study used a direct measure specific to the ego involvement motive. In one study, the operational induction of ego involvement was accomplished through

<sup>1</sup> The initial phase of this investigation was begun while the author was a postdoctoral fellow in the Psychology Department, Western Psychiatric Institute, University of Pittsburgh. Grateful acknowledgments are extended to Arthur R. Thomas and Robert E. Rankin of West Virginia University for their invaluable help in serving as experimenters for some of the studies and in their reading of the manuscript. Appreciation also is extended to Robert Eigenbrode for permission to use the performance items of the Rating Test, which he standardized as part of his work in a graduate psychology seminar.

both a preselection of subjects as well as the administration of appropriate instructions (instructions were high or low ego involving in relation to whether subjects were high in *n* Achievement or high in *n* Affiliation). And whereas overt performance on a task was barely affected by the instructions, a more direct measure of motivation, i.e., PGR, did show effects, albeit weak effects, from the ego involving instructions (Vogel et al., 1959). In another study, the use of a fairly direct measure of motivation (level of aspiration) showed that when subjects were forced to choose between two responses, clear-cut effects from ego involving instructions were obtained (Feather, 1959). A quite different conceptual approach that may permit a more thorough exploration of the role of ego involving instructions comes from the field-theoretical writings of Wertheimer (1959). He suggests that it is easier to remember relationships, or relational qualities, than absolute qualities of stimuli. Although he referred to memory and to problems of discrimination learning, this suggestion may be relevant to the present problem. If suitably generalized, the suggestion would be that people respond more easily to relational than absolute qualities; as applied to the present problem, one would predict that maximum effects from ego involving instructions will be obtained when subjects respond to the difference between two instructions, instead of responding to merely a single ego involvement instruction as in the more usual experimental procedure.

In line with various of the considerations raised above, the present investigation seeks to delimit some of the characteristics of the ego involvement variable so as to render it more useful for experimental research. Specifically, the present series of studies attempts to explore some important aspects of the stimulus characteristics of ego involving instructions, to tie these where possible to characteristics of ego involvement motivation, and in turn to relate these to a direct response measure of the ego involvement motivation. In this multifaceted aim, emphasis is placed on exploring the degree of, and conditions for, the effectiveness of ego involving instructions, and to delineate the valence

characteristics of the ego involvement motivation as these pertain both to the instructions and to the response measures employed in this investigation. Specifically, the following questions were examined.

1. Although several means are possible to test the effectiveness of ego involving instructions, one test for their effectiveness is to establish that ego involving instructions constitute a discriminable stimulus dimension. That is, the majority of subjects when given both the high and the low ego involving instructions should be able to make a reliable distinction between them.

2. Successful performance on a high ego involvement task is more rewarding, or has a higher valence, than on a low ego involvement task. This is an assumption generally held for the ego involvement motivation and at times has been made the operational basis for defining ego involving tasks (e.g., Bayton, 1943). But this assumption has never, to the author's knowledge, been tested directly. That is, no data appear to exist which show that a given high level of performance is more desirable on the high ego involvement task than on the low.<sup>2</sup> A slight extension of this hypothesis would also predict that subjects strive for a higher level of success on the high as against the low ego involvement task (i.e., set higher levels of aspiration for the high as against the low ego involvement task). Evidence for this has been obtained in a study with children (Feather, 1959).

3. Ego involving instructions are far more effective when subjects are in a stimulus-difference set. That is, when subjects have a set of contrasting one ego involvement level to another level, more clear-cut effects from ego involving instructions are obtained than when subjects perceive the instructions as merely an isolated, single event.

4. If under conditions of high ego involvement subjects have an increased desire to do well on the task, one direct measure of the

<sup>2</sup> Since level of aspiration is considered to be a function of both the valence and the probability of attainment of a given goal (Lewin, Dembo, Festinger, & Sears, 1944; Siegel, 1957), level of aspiration data do not provide a direct test of the valence assumption made in Prediction 2. The Transitivity Test devised for the present investigation is an attempt to arrive at such a direct test.



subjects' degree of ego involvement in the task would be "interest in doing well on a task." To establish the validity of this as a response measure of ego involvement ( $EI_R$ ), two essential criteria must be met: subjects should state higher "interest in doing well on a task" on the high than the low ego involvement task; and the degree of interest in doing well on an ego involving task should remain constant regardless of how well in fact the subject does perform (except at the extreme limits of performance), since this interest is defined by the nature of the task (i.e., by the potential rewards of success on the task) rather than by actual success. This latter criterion distinguishes "interest in doing well on a task" from level of aspiration as a direct response measure of ego involvement, since level of aspiration has been both hypothesized and empirically found to vary with actual performance level on a task (e.g., Lewin et al., 1944; Steisel & Cohen, 1951).

5. A relevant internal drive induced in conjunction with ego involving instructions increases the effect of the ego involving instructions.

## METHOD

To test these predictions, and to clarify the extent to which ego involving instructions are effective in eliciting response differences, one major response measure used in the following series of studies was level of aspiration (LA). The author's own experience with LA (Ferguson, 1958) indicated this to be a sensitive measure of motivation, and recent data (Feather, 1959) reversed earlier findings which had failed to show significant effects of ego involvement on level of aspiration. The wording used for LA in this series of studies was such that subjects stated *desired* rather than *anticipated* success, i.e., subjects stated level of aspiration (LA) and not level of expectation (LE). This was necessary to test the hypothesis concerning success preference for the High as against the Low Ego Involvement condition. In line with this, the response measure of ego involvement ( $EI_R$ ) explored in this investigation was "interest in doing well on the task" instead of "interest in doing the task." For the latter is presumably more contingent on actual than desired success, and also has no necessary relationship to an activity involving the subject's status.

The ego involvement instructions used in this investigation were selected to represent the more typical wording of the two extreme levels of ego involvement instructions most commonly found in the experimental literature. The purported difference between High and Low Ego Involvement instructions

is that the subject's status (self-esteem and/or social esteem) is at stake in the former type of task but not, or only minimally, in the latter type of task.

## Procedure

**Tasks.** Two main tasks were used in the present series of studies. On the one task, the Rating Test, a subject received only a single ego involvement instruction; on the other, the Transitivity Test, the subject received both ego involvement instructions. The studies varied with regard to whether one, or the other, or both tasks were given to the subjects. The rationale for administering the experimental tasks in any given study was dependent on the specific hypothesis being tested by that study. As will be seen, the findings reported in the subsequent section are clearly related to the specific task procedure employed.

**Rating Test.** This consisted of a small five-page booklet. On the first, third, and fifth pages, subjects had merely to make a rating from 0 to 50. On the second and fourth pages, subjects made no ratings but filled in performance items. The Rating Test thus consisted of three rating trials and two performance trials. Each performance trial contained 50 digit substitution items, patterned after the Wechsler Adult Intelligence Scale. Successful and Failing performance was predetermined by allotting subjects enough time to complete approximately 70% of the items per trial in the Success condition and approximately 40% of the items in the Failure condition.

In addition to a rating scale, the first page contained ego involvement instructions. A given subject received only the High or the Low instructions. High Ego Involvement instructions stated:

You are being asked to take a brief intelligence test. Results on this test give a very good indication of a person's intelligence and of the likelihood of career success.

The instructions for Low Ego Involvement were:

You are being asked to take a test which is in the process of being developed as part of a large experiment. Since you are the first group to do this task, individual scores cannot be adequately evaluated, but your performance will provide some valuable information on what the next steps should be in the further development of this task.

The Rating Test was used to obtain one of two responses. Level of aspiration (LA) was one, the other was an ego involvement response ( $EI_R$ ) designed especially for this investigation. For LA, the subject placed a check on a scale from 0 to 50 to indicate the number of items he would like to complete in the allotted time. For  $EI_R$ , the subject placed a check on a scale from 0 to 50 to indicate how interested he was in doing well on the task. Each subject made only LA or  $EI_R$  ratings. The ratings on the first page were made before performance was begun, only the ego involving instructions being provided as basis for the subjects' ratings. Ratings on



the third page, following the first performance trial, referred to the subsequent performance trial: how many items the subject would like to complete on the remainder of the task (LA) or how interested the subject was in doing well on the remainder of the task (EI<sub>R</sub>). Ratings on the fifth page (LA or EI<sub>R</sub>) referred to another task like the one the subject had just completed but which he was not going to be asked to do at the present time.

**Transitivity Test.** This consisted of one sheet that contained both ego involvement instructions already described, plus 15 pairs of items. Unlike the Rating Test which provided the subject with only one ego involvement instruction, the Transitivity Test forced the subject to contrast one instruction with another. This test was designed not only to ascertain the subjects' responses in a stimulus-difference situation, but also to test the transitivity and direction of the subjects' success preferences for the two ego involvement conditions. (Transitivity exists when A is preferred to B and B is preferred to C, and A in turn is preferred to C.)

On the top of the sheet were two paragraphs, each paragraph giving the same ego involvement instructions provided on the Rating Test. For control, half the sheets had the High Ego Involvement instructions for the first paragraph, the other half had the Low Ego Involvement instructions for the first paragraph. The High Ego Involvement task was labeled Intelligence Test and the Low Ego Involvement task was labeled Experimental Test. Subjects then had to make 15 pair comparisons, made up of six units: three goals (50, 40, and 30 items completed on a task consisting of 50 items), for the two ego involvement tasks. Each unit was compared once to every other unit, and subjects had to check the preferred item for each of the 15 pairs of items. An illustrative example was provided:

- (a) 30 on Experimental Test \_\_\_\_\_  
 (b) 30 on Intelligence Test \_\_\_\_\_

If you have checked (a), that means you would prefer to complete 30 items on the Experimental Test in contrast to 30 on the Intelligence Test.

### Subjects

Subjects were drawn from three institutions: University of Pittsburgh (introductory sociology and introductory psychology), West Virginia University (introductory psychology), and Pennsylvania State University (first year student nurses who were in a 3-year hospital program). Subjects took the experiments in their regular class time, without prior knowledge that they were going to be subjects, i.e., they were a captive audience and not volunteers. The series of studies covered a 2-year period, but no subject served in more than one study. Assignment of subjects to the experimental conditions within each study was such that any effects of the experimental variables were not due to systematic subject bias. Tests of significance were made only between subjects serving in each self-contained study.

Where the  $N$  within cells was not equal, a non-parametric analysis of variance which yields chi square (Wilson, 1956) was used; where  $N$  within cells was equal, an  $F$  test was used to test the significance of the data.

### RESULTS

The data indicate several clear-cut findings:

1. Ego involvement instructions are effective but weak; i.e., they yield clear-cut effects, but only under certain conditions. And though they are reliably distinguished by a majority of subjects, they fail to be so distinguished by a sizable minority of subjects.

2. The valence of success is greater for High than Low Ego Involvement tasks.

3. Instructions yield clear-cut and consistent effects only when subjects define the ego involvement conditions in a context of contrast.

4. Even with the addition of a relevant internal drive, ego involving instructions fail to produce clear-cut and consistent effects when subjects are not in a stimulus-difference set (i.e., when not defining the ego involvement conditions in a context of contrast).

5. Interest in doing well on a task (EI<sub>R</sub>) is a valid and reliable measure of ego involvement, but it is sensitive to ego involving instructions only for selected subjects.

Evidence for the effectiveness of ego involvement instructions comes from two main sources: on the Transitivity Test the majority of subjects were able to make consistent goal preference discriminations between tasks differing in degree of ego involvement; when subjects were in the appropriate set, their responses on the Rating Test showed significant differences between High and Low Ego Involvement conditions. Four groups, or a total of 197 subjects, took the Transitivity Test (Table 1). In each group, the majority of subjects, ranging from 54% to 69%, made a reliable goal preference distinction between the High and Low Ego Involvement tasks. The nature of this distinction was a highly significant preference for success on the High as against the Low Ego Involvement task. The vast majority of subjects who made transitive choices did so in the predicted direction, i.e., they chose 50 on an intelligence test as their most preferred goal, with 40 on

TABLE 1

CONSISTENCY OF PREFERENCES FOR PERFORMANCE GOALS ON AN INTELLIGENCE TEST AS COMPARED TO AN EXPERIMENTAL TASK

	Student nurses (N = 33)	University of Pittsburgh (N = 49)	University of Pittsburgh (N = 53 <sup>a</sup> )	West Virginia University (N = 62)
Number of subjects making transitive choices	18	34	34	41
Number of subjects making transitive choices in the predicted direction <sup>b</sup>	14	24	26	31

<sup>a</sup> Although the total N in this group was 54, the Transitivity Test was not handed in by one subject in that group.  
<sup>b</sup> With six goal units in any possible combination of ranks, the chance probability of getting the combination of "50 on the Intelligence Test" in first rank and "40 on the Intelligence Test" in either second or third rank is 1 in 15 (48/720). Using the standard error of a proportion (where the obtained proportion is, e.g., 14 out of 18, and the chance proportion is 1 out of 15), the number of subjects whose transitive choices are in the predicted direction far exceeds the .01 level of significance for each of the four groups.

an intelligence test as their second or third preference.<sup>3</sup> Clearly, when ego involvement instructions constitute a discriminable dimension for subjects, their preference is very strongly in favor of success on the High Ego Involvement task. This same preference direction was seen in the subjects' responses on the Rating Test. When clear-cut ego involvement effects were obtained, the LA and EI<sub>1</sub> ratings were significantly higher (higher goals desired, and greater interest in doing well) for the High than the Low Ego Involvement conditions.

Of interest is the fact, however, that ego involvement effects *failed* to occur unless subjects were in the appropriate set. Only when subjects were aware of a contrast between two ego involvement stimuli did their responses show a clear-cut effect from the instructions. It will be recalled that on the Transitivity Test subjects were responding concurrently to two ego involvement instructions, but on the Rating Test they were exposed to only one specific instruction. When the Rating Test *followed* the Transitivity Test, subjects' ratings showed significant differences between the two ego involvement conditions (Table 2). When the Rating Test *preceded* the Transitivity Test, no significant effects from the ego involvement conditions were obtained, and this was true even for those subjects who had a clear-cut preference in favor of success on the High Ego Involvement task (Table 3). Thus in spite of subjects' inherent preference for success on the one as against the other task, without an appropriate set subjects' goal setting behavior fails to reflect this preference. This points to the conclusion that without an appropriate set, subjects' responses were not based primarily on the ego involving instructions but on some other stimulus variables outside of the experimenter's control and manipulation.

The evidence thus far shows that ego involvement instructions do form a discriminable dimension, at least for a majority of subjects, and this dimension has a valence scale such that success is preferred on the High Ego Involvement task. But ego involvement instructions are not strong stimuli. This is seen not only by the large number of subjects who failed to make transitive choices between the two ego involvement tasks, but also by the fact that when subjects were given only a single ego involvement condition without an appropriate set, consistent and clear-cut ego involvement effects failed to appear. This was found so unvaryingly that it cannot be dismissed as merely "negative effects." Two studies were conducted without the use of the Transitivity Test. In the first study, subjects took the experiment early in the semester and the experimenter (a woman) was a stranger to the subjects; in the second study the experiment was conducted one week before final exams and the

involvement effects *failed* to occur unless subjects were in the appropriate set. Only when subjects were aware of a contrast between two ego involvement stimuli did their responses show a clear-cut effect from the instructions. It will be recalled that on the Transitivity Test subjects were responding concurrently to two ego involvement instructions, but on the Rating Test they were exposed to only one specific instruction. When the Rating Test *followed* the Transitivity Test, subjects' ratings showed significant differences between the two ego involvement conditions (Table 2). When the Rating Test *preceded* the Transitivity Test, no significant effects from the ego involvement conditions were obtained, and this was true even for those subjects who had a clear-cut preference in favor of success on the High Ego Involvement task (Table 3). Thus in spite of subjects' inherent preference for success on the one as against the other task, without an appropriate set subjects' goal setting behavior fails to reflect this preference. This points to the conclusion that without an appropriate set, subjects' responses were not based primarily on the ego involving instructions but on some other stimulus variables outside of the experimenter's control and manipulation.



TABLE 2  
EFFECTS OF EGO INVOLVEMENT (EI) ON LEVEL OF ASPIRATION (LA) AS A  
FUNCTION OF SUBJECTS' SET

Trial	When LA precedes the Transitivity Test (University of Pittsburgh <sup>a</sup> )					When LA follows the Transitivity Test (University of Pittsburgh <sup>a</sup> )				
	High EI (N cell = 22)		Low EI (N cell = 27)		$\chi^2$	High EI (N cell = 27)		Low EI (N cell = 27)		$\chi^2$
	M	SD	M	SD		M	SD	M	SD	
First (Preperformance)	43.9	9.3	42.9	9.8	.14*	45.3	8.4	38.1	8.2	16.71**
Second (After Failure)	39.6	9.3	36.6	12.5		37.1	12.9	28.3	11.5	
Third (After Failure)	39.6	8.9	30.8	16.9		36.4	13.3	26.1	12.7	
Between High-Low EI										

\* Although a significant difference between the two levels of ego involvement exists for the third trial ( $t = 2.33$ ,  $p < .05$ ), the combined results for all three trials show no significant effect from the Ego Involvement variable.

\*\*  $p < .01$ . The chi square is based on Wilson's (1956) nonparametric analysis of variance, with 1  $df$  for a High vs. Low EI comparison.

experimenter (a man) was the subjects' own instructor. Both studies showed that when subjects were not in the appropriate set, effects of ego involvement instructions failed to be significant even though subjects responded to Success and Failure as predicted (Tables 4 and 5). That is, LA was significantly lower under conditions of Failure, and  $EI_R$ , as hypothesized, was stable with performance. The lack of clear-cut LA and  $EI_R$  differences for ego involvement conditions must be assumed to be due to the ineffectiveness of the ego involvement instructions, since LA was extremely sensitive to performance and  $EI_R$  was highly reliable. (For the groups cited in Table 4, four correlations for  $EI_R$

were computed, for first vs. second trial and first vs. third trial on the combined Success and Failure subjects. These correlations ranged from .77 [ $N = 63$ ] to .89.)

It could be argued that the lack of effects from instructions, shown in Tables 4 and 5, is due to the fact that the High Ego Involvement instructions were not ego involving enough and that the Low Ego Involvement instructions were too ego involving (i.e., not low enough in ego involvement). To test this, two experiments were conducted so that ego involvement would be both raised and lowered. An initial study explored not only the role of a relevant drive (subjects taking the experiment immediately prior to a course ex-

TABLE 3  
EGO INVOLVEMENT EFFECTS AS A FUNCTION OF SET AND PREFERENCE<sup>a</sup>

Rating	When ratings preceded the Transitivity Test					When ratings followed the Transitivity Test				
	LA (University of Pittsburgh <sup>a</sup> )				$\chi^2$	LA (University of Pittsburgh <sup>a</sup> )				$\chi^2$
	High EI (N = 5)		Low EI (N = 12)			High EI (N = 8)		Low EI (N = 11)		
	M	SD	M	SD		M	SD	M	SD	
First rating	46.2	8.5	48.2	3.4	.26	49.8	.7	40.9	7.7	12.45**
Second rating	39.0	10.2	40.2	13.3		41.9	8.4	25.9	12.0	
Third rating	38.5	10.6	35.1	17.2		38.1	11.9	25.8	13.6	
Between High-Low EI										

<sup>a</sup> When subjects' transitive choices clearly favored success on the High EI task. Because only a very limited number of subjects strongly favored success on a Low Ego Involvement task, the present comparisons are based only on ratings of subjects who strongly favored success on a High Ego Involvement task, i.e., subjects whose transitive choices were "50 and 40 on an Intelligence Test" as two of the top three ranks and "50 on an Experimental Test" as no higher than third rank.

\*\*  $p < .01$ .



TABLE 4  
EFFECTS OF EGO INVOLVEMENT AND SUCCESS-FAILURE ON EI<sub>R</sub> AND LA

Trial	EIR						$\chi^2$	LA						$\chi^2$
	High EI			Low EI				High EI			Low EI			
	N	M	SD	N	M	SD		N	M	SD	N	M	SD	
First														
Success	22	36.9	12.1	32	38.5	10.3		36	40.2	9.2	38	36.4	12.3	
Failure	32	41.8	6.7	31	40.3	11.4		32	37.5	10.1	29	36.1	9.5	
Second														
Success		38.3	12.7		39.4	11.0			42.3	8.4		39.1	10.2	
Failure		42.4	10.1		43.2	11.0			29.8	9.5		26.8	10.3	
Third														
Success		36.1	15.3		38.6	11.5			41.8	8.4		39.4	9.1	
Failure		43.6	7.5		41.7	11.8			25.5	11.8		22.5	11.9	
First														
Between High-Low							.01							.60
Between Success-Failure							1.85							.14
Second														
Between High-Low							.30							2.13
Between Success-Failure							1.67							35.00**
Third														
Between High-Low							.24							.07
Between Success-Failure							3.88*							55.98**

<sup>a</sup> Although  $p < .05$ , the effect of performance on EI<sub>R</sub> fails to be significant when the Success-Failure differences on the first (preperformance) trial are taken into account. When the change scores (from the first to the third trial) of the two combined Success groups were compared with those of the combined Failure groups, the obtained  $t$  of 1.35 fell short of the 2.00 value required for a .05 level of significance (with 62 and 53  $df$  and heterogeneous variances). Thus the hypothesis is supported that differential success does not have a significant effect on EI<sub>R</sub>.

\*\*  $p < .01$  (chi square at the .01 level is 6.64).

amination vs. taking the experiment in a relaxed and informal atmosphere) but also whether anonymity under low ego involvement would help produce the expected ego involvement effects. If these two added variables would be effective in eliciting significant differences from the ego involving instructions, the predicted differences would be

TABLE 5  
EFFECTS OF EGO INVOLVEMENT AND FAILURE WHEN SUBJECTS<sup>a</sup> ARE IN AN EGO INVOLVED RELATIONSHIP WITH THE EXPERIMENTER  
( $N$  per cell = 26)

E <sub>1</sub> INVOLVED RELATIONSHIP WITH THE EXPERIMENTER (N per cell = 26)										
Trial	EIR				F	LA				F
	High EI		Low EI			High EI		Low EI		
	M	SD	M	SD		M	SD	M	SD	
First (Preperformance)	39.8	6.0	39.1	7.9		42.0	9.5	43.4	6.3	
Second (Failure)	38.8	8.5	41.0	8.0		36.0	9.7	32.4	11.4	
Third (Failure)	37.5	11.2	39.4	10.4		33.4	13.3	28.4	13.7	
First vs. Second					.14					.25
Between High-Low EI					.27					41.18**
Between Neutral-Failure										
First vs. Third					.07					.49
Between High-Low EI					.58					42.33**
Between Neutral-Failure										

<sup>a</sup> Subjects are from West Virginia University.  
\*\*  $p < .01$ . Using the analysis of variance design suggested by Edwards (1950) for repeated measurements of the same subjects, the  $F$  ratios represent 1, 50  $df$ . ( $F$  at the .01 level is 7.17.)

TABLE 6  
EFFECTS OF EGO INVOLVEMENT ON LA WHEN A RELEVANT DRIVE AND ANONYMITY  
UNDER LOW EI ARE ADDITIONAL VARIABLES

Trial	First experiment (Student nurses)				F	Second experiment (West Virginia University)				x <sup>2</sup>
	High Drive		Low Drive			High Drive		Low Drive		
	(N cell = 12)		(N cell = 12)			(N cell = 21)		(N cell = 20)		
	M	SD	M	SD		M	SD	M	SD	
First (Preperformance)										
High EI	43.1	7.2	39.6	9.3		40.8	9.5	46.6	5.4	
Low EI	40.9	8.8	36.5	7.4						
Low EI-Anonymity	41.9	6.5	38.2	8.2	.58	46.2	7.9	38.6	12.0	.78
Between EI conditions										
Second (Failure)										
High EI	35.2	11.1	39.6	9.0		34.7	8.4	42.2	10.6	
Low EI	32.0	10.3	32.1	8.9						
Low EI-Anonymity	28.2	7.3	33.5	5.2	3.34	37.9	11.8	29.9	12.4	3.12
Between EI conditions										
Third (Failure)										
High EI	38.1	11.1	43.9	6.7		34.9	11.7	42.9	11.3	
Low EI	30.0	11.2	30.1	10.7						
Low EI-Anonymity	25.8	4.2	32.5	9.8	10.98**	37.1	15.2	28.8	12.1	3.96*
Between EI conditions										

\*  $p < .05$ . Though instruction effects occur only on the third trial, Instruction  $\times$  Drive interaction is significant for all three trials, in opposite direction to that predicted.  
\*\*  $p < .01$ .

in the direction of highest LA for the High Drive-High Ego Involvement condition and lowest LA for the Low Drive-Low Ego Involvement-Anonymity condition, with the remaining four conditions yielding appropriate LAs between these two extremes. The predicted effects were not obtained, in either the initial or the follow-up study. The latter used only four conditions (Low Ego Involvement

being given only under anonymity) to maximize the possibility of obtaining the predicted ego involvement effects within the limit of the available number of subjects. In both studies, ego involvement effects appeared to a significant extent only on the third trial, and then not as predicted (Table 6). Drive effects were not significant in either study, and analysis of variance between the

TABLE 7  
EFFECT OF EGO INVOLVEMENT ON EI<sub>R</sub> WHEN SUBJECTS<sup>a</sup> ARE IN AN APPROPRIATE SET

Trial	For total group				$F_c$	For selected subjects <sup>b</sup>				$\chi^2$
	High EI		Low EI			High EI		Low EI		
	(N cell = 31)		(N cell = 31)			(N cell = 12)		(N cell = 7)		
	M	SD	M	SD		M	SD	M	SD	
First (Preperformance)	44.1	6.4	40.1	10.4	2.36 .37	44.8	5.7	36.4	11.3	8.52** .14
Second (Failure)	44.2	6.8	41.6	11.8		45.5	4.6	36.6	13.7	
Third (Failure)	44.1	7.8	41.7	10.0		44.6	6.9	34.7	13.9	
Between High-Low EI										
Between trials										

<sup>a</sup> Subjects are from West Virginia University.  
<sup>b</sup> These subjects strongly preferred success on a High EI task (see footnote to Table 3).  
<sup>c</sup> The F tests follows Edwards' (1950) suggested analysis for repeated measurements on the same subjects.  
\*\*  $p < .01$ .

two Low EI conditions in the initial study showed no significant effects for Anonymity.<sup>4</sup> In the second experiment, Drive and Instruction interaction was significant (it was not in the first study), but in the opposite direction from that predicted. These data serve only to confirm the results from the preceding series of studies, and indicate that when subjects are not in an appropriate set, clear-cut and consistent effects from ego involvement instructions fail to occur. Moreover, it does not appear tenable to account for the lack of ego involvement effects by an insistence that the fault lies in the particular instructions used in this series of studies. For not only do these instructions represent the type often used in the experimental literature, but even when the High instructions were augmented by an ego involvement drive and the Low instructions were made minimally ego involving through subjects' performance and ratings being anonymous, the instructions failed to be reliably effective.

One remaining study was carried out. Since a method had been found for eliciting instruction effects, by giving subjects the Transitivity Test first and thereby inducing an appropriate set, a means was now available to test whether  $EI_R$  is a suitable response measure for ego involvement. As seen in Table 7, subjects whose preference was clearly in favor of success on the High Ego Involvement task showed a significant difference between the High and Low Ego Involvement conditions. But when the ratings of all subjects are compared, significant ego involvement effects were not obtained even though the Transitivity Test preceded the Rating Test. Thus  $EI_R$  is reliable, and a valid measure of ego involvement for selected subjects; but for group data it is not as sensitive to ego involving instructions as is LA. The use of  $EI_R$  instead of LA would be preferable, however, in cases where stability of response is desired, since LA does vary with performance and

<sup>4</sup> Since the conditions of High vs. Low Drive and Anonymity vs. Nonanonymity were introduced into the experiments to heighten and lower ego involvement motivation, they can be viewed as additional categories of  $EI_s$ . The negative results obtained from these two studies therefore suggest that the conclusions derived from the studies using only instructions as  $EI_s$  may have wider generality.

$EI_R$ , as hypothesized, does not do so even when it is sensitive to the ego involving conditions (see chi square for Between trials in Table 7).

## DISCUSSION

The data that have been presented show clearly that not only are ego involvement instructions less potent than has been generally assumed, but also that they will fail to produce clear and consistent response differences unless subjects define the stimuli as part of a discriminable dimension. Evidence for the weakness of ego involving instructions comes from several independent sets of data, but major evidence for it is the fact that a large minority of subjects do not discriminate reliably between two levels of ego involvement, in terms of the implicit valence scale generally assumed to exist for this variable. An argument against the findings of the present investigation might be raised on the basis that other published studies show clearly significant effects from ego involving instructions without the use of an appropriate prior set. However, the review of the literature already presented should leave little doubt that the evidence regarding the effectiveness of ego involvement instructions is far from consistent. In view of publication policy regarding negative findings, one can only surmise that other investigations which failed to obtain significant ego involvement effects are not lacking. Even among published studies, negative results and data in the direction opposite to that predicted are reported. Since ego involvement has been shown in this investigation to be a reliable phenomenon, even though weak, one can expect that occasional significant results will be obtained without prior set induction.<sup>5</sup>

The findings of the present investigation not only shed new light on the ego involvement variable, but have implications for other motivational variables as well. The data from the present series of studies appear to make the most sense if one assumes that the human

<sup>5</sup> Especially the forced-choice technique used by Feather (1959) would be a likely means of obtaining effects from ego involving instructions, since this at least narrows the response alternatives even if it does not narrow the stimulus field.



adult in most situations is exposed to a great number of internal and external stimuli at any given moment. When, e.g., a single ego involvement condition is presented, subjects are probably exposed to many additional motivations: the need to achieve, the need to attend to the task, the desire to please, the avoidance of failure, and so on. Other stimuli, aside from motivational ones, are also present. (A recent study—Underwood & Postman, 1960—has explored this question in another context.) If the experimental stimulus is already a weak one, as has been demonstrated for ego involving instructions in the present series of studies, the probability that subjects will respond to the experimental stimulus and not to other ones is quite small. However, when an appropriate set is induced, there is both a narrowing of the stimulus field and a specifying of the unique properties of the experimental stimulus. Once the subject is in this set, the probability of his responding to the *unique* stimulus that the experimenter expects him to respond to is greatly increased. In the present investigation, the Transitivity Test was a simple way of inducing such a stimulus set. Without it, subjects failed to respond to the experimental stimulus in terms of its unique properties even though they were able to ascribe these properties to the stimulus. The most striking evidence for this came from the data of the subjects whose choices were transitive and in favor of success on a High Ego Involvement task. Without an appropriate set, their level of aspiration was uniformly high, regardless of the task instructions. This suggests that they were responding to other variables, rather than primarily to the experimental one. But when subjects were in an appropriate set, these same instructions produced responses far more in line with the main experimental condition, the most notable effect being a lowered LA for the Low Ego Involvement condition. This finding of the importance of a relevant stimulus set suggests that possibly other motivational variables could be brought into sharper focus by an appropriate set induction. Studies in other areas of human motivation (e.g., stress, anxiety, n Achievement) have suffered some of the same inconclusive results found in the literature of

ego involvement. When motivation in human subjects is experimentally induced, this may indeed elicit only weak stimuli, as critics of this approach to human motivation have been free to point out. Granting the possible limits of laboratory defined motivations, the induction of a relevant set comparable to that employed in the present investigation may make for more effective study of such motivations.

#### SUMMARY

The present investigation explored some important aspects of the ego involvement variable so as to render it more useful for experimental research. A series of studies revealed the following:

1. Ego involvement instructions are effective but less potent than has been generally assumed in eliciting the desired motivation.
2. To a very significant degree, subjects prefer success on a High as against a Low Ego Involvement task.
3. Consistent and clear-cut effects from ego involvement instructions occur only when subjects are in an appropriate set, of perceiving a given instruction as part of a dimension consisting of contrasting levels of ego involvement. Without such a set, ego involvement instructions are not reliably effective, and this is true even when instructions are combined with a relevant drive and when subjects under low ego involvement take the task anonymously.
4. Interest in doing well on a task is a valid response measure of ego involvement for selected subjects. As hypothesized, it is stable with performance. But for unselected subjects it is not as sensitive to ego involvement instructions as is level of aspiration which is *not* stable with performance.

Limiting the number of cues to which subjects respond in an experiment dealing with motivation appears to be a necessary condition for eliciting significant response differences. The Transitivity Test devised in this investigation contrasted two levels of ego involvement and thus focused the subjects on the specific experimental cues. Through such a technique, which emphasizes relations rather than absolute attributes of experimental instructions, it should be possible to establish relevant orientations for other motivational

variables and thus contribute to added precision in their study.

# REFERENCES

- ALPER, THELMA G. Task-orientation vs. ego-orientation in learning and retention. *Amer. J. Psychol.*, 1946, 59, 236-248.
- ALPER, THELMA G. Predicting the direction of selective recall: Its relation to ego strength and *N* achievement. *J. abnorm. soc. Psychol.*, 1957, 55, 149-165.
- BAYTON, J. A. Interrelations between levels of aspiration, performance, and estimates of past performance. *J. exp. Psychol.*, 1943, 33, 1-21.
- CASTANEDA, A., & PALERMO, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, 50, 175-179.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
- FEATHER, N. T. Subjective probability and decision under uncertainty. *Psychol. Rev.*, 1959, 66, 150-164.
- FERGUSON, EVA D. The effect of sibling competition and alliance on level of aspiration, expectation, and performance. *J. abnorm. soc. Psychol.*, 1958, 56, 213-222.
- HEVER, A. W., & O'KELLY, L. I. Studies in motivation and retention: II. Retention of nonsense syllables learned under different degrees of motivation. *J. Psychol.*, 1949, 27, 143-152.
- HOLT, R. R. Effects of ego-involvement upon levels of aspiration. *Psychiatry*, 1945, 8, 299-317.
- IRWIN, R. W., & MINTZER, MARCIA G. Effect of differences in instructions and motivation upon measures of the level of aspiration. *Amer. J. Psychol.*, 1942, 55, 400-406.
- KAUSLER, D. H. A study of the relationship between ego-involvement and learning. *J. Psychol.*, 1951, 32, 225-230.
- LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. I. New York: Ronald, 1944. Pp. 333-378.
- ROMANOW, CONCETTA V. Anxiety level and ego involvement as factors in concept formation. *J. exp. Psychol.*, 1958, 56, 166-173.
- RUSSELL, W. A. Retention of verbal material as a function of motivating instructions and experimentally-induced failure. *J. exp. Psychol.*, 1952, 43, 207-216.
- SHEDD, C. L., & ANGELINO, H. R. The effect of ego-involvement on learning. *Proc. Okla. Acad. Sci.*, 1952, 33, 285-288.
- SIEGEL, S. Level of aspiration and decision making. *Psychol. Rev.*, 1957, 64, 253-262.
- STEISEL, I. M., & COHEN, B. D. The effects of two degrees of failure on level of aspiration and performance. *J. abnorm. soc. Psychol.*, 1951, 46, 79-82.
- UNDERWOOD, B. J., & POSTMAN, L. Extraexperimental sources of interference in forgetting. *Psychol. Rev.*, 1960, 67, 73-95.
- VOGEL, W., RAYMOND, SUSAN, & LAZARUS, R. S. Intrinsic motivation and psychological stress. *J. abnorm. soc. Psychol.*, 1959, 58, 225-233.
- WERTHEIMER, M. On discrimination experiments: I. Two logical structures. *Psychol. Rev.*, 1959, 66, 252-266.
- WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, 53, 96-101.

(Received March 6, 1961)

## PARENTAL ATTITUDES ASSOCIATED WITH SOCIAL DEVIANCE IN PREADOLESCENT BOYS<sup>1</sup>

C. L. WINDER<sup>2</sup> AND LUCY RAU

*Stanford University*

This is a study of the relationships between parent attitudes and social deviance of preadolescent boys. "Social deviance" is here defined as the extent to which a child's characteristic interpersonal behavior is extreme on one or more of several general dimensions as reported by his peers.

### METHOD

#### *Measures*

The measure used for social deviance is the Peer Nomination Inventory (PNI). This is a modified "guess who" technique, which is administered to classroom groups. Pupils are asked to nominate members of the class on each of 64 items. Each item refers to a fairly specific aspect of behavior and is phrased in language characteristically used by children of this age in describing each other. The instrument yields a score for each child on five variables: Aggression, Dependency, Withdrawal, Depression, and Likeability. The psychometric properties of the PNI are described by Winder and Wiggins (1960). It should be noted that all the dimensions are intercorrelated (see Table 1). Thus the instrument may be considered a measure of general social deviance which is negatively related to "likeability," or sociometric choice.<sup>3</sup>

We assume that parent characteristics are important antecedents of social deviance in boys of this age. Accordingly we have developed a questionnaire measure of parent attitudes, the Stanford Parent Attitude Questionnaire. The current version of the questionnaire has two forms, one for mothers and one for fathers. They consist of 491 and 518 items, respectively, grouped into 27 and 28 scales. The list of scales, including six summary scales, and the number of items for each are given in Table 2. The ma-

jority of the items, except for those in the Parental Adjustment and the Models groups of scales, are identical for mothers and fathers. One scale, Closeness to father, applies to fathers only.<sup>4</sup>

The variables to be measured were selected on the basis of the available studies of parental antecedents of personality development in children. These studies include those of child rearing practices of parents of normal preschool and older children, e.g., Sears, Maccoby, and Levin (1957), and Miller and Swanson (1960). Other researchers, e.g., Bandura (1960), have systematically studied parents of particular groups of deviant children. In addition, our choice of relevant parent characteristics was influenced by clinical experience.

Items were selected on an a priori basis to represent each of the variables. The main sources of items were verbatim transcriptions of interviews with parents of deviant and normal children.<sup>5</sup> These items samples reported actions toward child and others, attitudes toward child, spouse, and self, and beliefs about child rearing and family roles and relationships. Whenever possible the original phrasing of the statements was preserved. The items are thus in the language actually used by parents in reference to particular people and situations, and so differ markedly from the kinds of items used in other parent questionnaires. Items were assigned to scales on the basis of the judgment of the authors and a graduate student in clinical psychology.<sup>6</sup>

The general strategy of the study was to collect responses to the parent questionnaire from a sample of mothers and fathers whose sons had received high, medium, or low scores on the PNI variables, Aggression and Dependency.

#### *Procedure*

**Sample.** The subjects of this study were drawn from a total population of 710 fourth, fifth, and

<sup>1</sup> The study was supported in part by USPHS Grants M-2745 and M-2745 (C1).

The cooperation and assistance of personnel of the Palo Alto Unified School District is gratefully acknowledged. Henry Gunn, M. John Rand, John Caffrey, and H. B. Gelatt were most helpful. Principals and teachers at nine elementary schools gave time and support which was essential.

<sup>2</sup> Now at Michigan State University.

<sup>3</sup> Although the items of the PNI are worded in the language of children of this age, the constructs which the scales purport to measure are in the authors' opinion central to deviant personality development. For the rationale of scale construction see Winder and Wiggins (1960). Validation studies are in progress.

<sup>4</sup> For each scale, the item having the highest correlation with the total scale score has been identified. The list of these items has been deposited with the American Documentation Institute. Order Document No. 7091 from ADI Auxiliary Publications Project; Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

<sup>5</sup> Interviews from studies by Bandura, by Sears and his associates, and from the Stanford Psychological Clinic were used.

<sup>6</sup> Ronald Cowan.



sixth grade boys in nine Palo Alto schools to whom the PNI had been administered. Cases were selected on the basis of their scores on either the Aggression variable or the Dependency variable. Half the cases were drawn equally from the top, middle, and bottom thirds of the distribution of scores on Aggression; the other half similarly represented the Dependency distribution. Letters were sent to the parents of these children requesting their participation in the study. Of the 262 families contacted, 64% responded, 55% agreed to take the questionnaire, and usable questionnaires were actually obtained from 108 fathers and 118 mothers. Comparisons were made to discover if the sample which took the questionnaire differed either from the total sample or from the sample which was asked to take the questionnaire but did not do so. There are no significant differ-

TABLE 1

SCALE RELIABILITIES AND INTERCORRELATIONS FOR THE PEER NOMINATIONS INVENTORY (PNI)

Scale	A	B	C	D	L
A. Aggression	.85				
B. Dependency	.74	.84			
C. Withdrawal	.29	.47	.88		
D. Depression	.48	.62	.62	.77	
L. Likeability	-.23	-.33	-.54	-.41	.94

Note.— $N = 710$ . All correlations significant at  $p < .001$ .

ences on the PNI variables except on Likeability. The critical ratio for the difference between the mean scores on Likeability for the respondents and the

TABLE 2

STANFORD PARENT ATTITUDE QUESTIONNAIRE VARIABLES, NUMBERS OF ITEMS, AND SCALE RELIABILITIES

Scale	Mothers		Fathers	
	Number of items	Reliability	Number of items	Reliability
I. Ambivalence				
1. Affection demonstrated	37	.66	37	.74
2. Contingent reward	10	.57	10	.50
3. Rejection	39	.66	40	.68
Total	86	.63	87	.67
II. Strictness				
4. Sex anxiety	15	.72	15	.71
5. Aggression anxiety	18	.65	18	.72
6. Restrictiveness	18	.57	18	.56
7. Demands for conformity	21	.62	21	.71
8. Consistency	7	.56	7	.52
9. Conscience	23	.61	23	.62
10. Deprivation of privileges	5	.50	5	.53
Total	107	.86	107	.89
III. Aggression and Punitiveness				
11. Demands for aggression	35	.73	36	.69
12. Punitiveness and physical punishment	11	.42	11	.61
13. Parental aggression	15	.65	15	.47
Total	61	.73	62	.70
IV. Parental Adjustment				
14. Self-esteem	20	.71	24	.68
15. Sociability	11	.52	11	.61
16. Positive father-mother relation	32	.77	46	.73
17. Inconsistency	14	.49	12	.29
Total	77	.67	93	.72
V. Models				
18. Use of models	9	.55	9	.70
19. Sex role expectations	26	.60	23	.45
20. Resemblance to parent	6	.42	10	.49
21. Authority	9	.48	7	.30
22. Closeness to father (fathers only)	0	—	9	.57
Total	50	.48	58	.61
VI. Mastery				
23. Democracy	21	.58	22	.62
24. Nonrestrictiveness	12	.32	12	.37
25. Rewarding independence	25	.69	25	.65
26. Child's mastery	15	.25	15	.40
27. Achievement standards	14	.74	14	.72
28. Reasoning	23	.71	23	.61
Total	110	.87	111	.85

TABLE 3  
EDUCATIONAL LEVELS OF PARENTS

Educational level	% Fathers	% Mothers
PhD or MD	7.3	0.0
MA, MS, or graduate work	23.6	7.2
AB or BS	24.5	23.0
Some college	15.5	34.9
High school + technical training	0.9	5.6
High school graduate	19.1	20.6
Less than high school graduate	9.1	5.6

"no shows" is 1.96 ( $p = .05$ ), the boys of the respondents being more popular. The means on the four deviance scales are all slightly higher for the "no shows," but not significantly so. In general we may say that the sample of families where parents took the questionnaire, as judged on the basis of social reputation of the boys, is not severely biased,

but does contain a slightly higher proportion of "well-adjusted" boys.

*Administration.* Questionnaires were administered to groups of parents on several evenings, using the meeting room facilities of the schools. It took most parents about 1.5 hours to complete the questionnaire, an arduous task for a good many parents.

We do not have detailed information on the demographic characteristics of the sample. Palo Alto is a predominantly middle- and upper middle-class suburban community. The educational level distribution of our subjects is shown in Table 3.

*Scoring procedure.* Each item was answered by checking one of four alternatives: Strongly Agree, Agree, Disagree, Strongly Disagree. A dichotomous method of scoring was used, based on the distribution of responses to each item. A plus score on a given item always indicated presence of the characteristic presumably measured by the scale. Either agreement or disagreement might be scored plus for a given item; a few items were dichotomized for scoring between the Strongly Agree or Strongly Disagree category and all others. Total score for each variable is simply the number of items answered in the scored direction.

TABLE 4  
CASES IN HIGH, MIDDLE, AND LOW CATEGORIES ON AGGRESSION AND DEPENDENCY: DISTRIBUTION ON OTHER PNI VARIABLES

Variable A (Aggression)		Variable B (Dependency)			Variable C (Withdrawal)			Variable D (Depression)		
		H	M	L	H	M	L	H	M	L
H	N	28	8	0	19	8	9	23	9	4
	%	78	22	0	53	22	25	64	25	11
M	N	8	18	9	6	16	13	8	16	11
	%	23	51	26	17	46	37	23	46	31
L	N	0	8	25	10	10	13	5	10	18
	%	0	24	76	30	30	40	15	30	55
Variable B (Dependency)		Variable A (Aggression)			Variable C (Withdrawal)			Variable D (Depression)		
		H	M	L	H	M	L	H	M	L
H	N	28	8	0	20	7	9	22	9	5
	%	78	22	0	56	19	25	61	25	14
M	N	8	17	9	9	14	11	12	16	6
	%	24	50	26	26	41	32	35	47	18
L	N	0	8	26	6	14	14	2	10	22
	%	0	24	76	18	41	41	6	29	65

Note.—Ranges of scores for H, M, and L groups were as follows: Variable A: H = 119 and up, M = 33–118, L = 0–32; Variable B: H = 94 and up, M = 33–93, L = 0–32; Variable C: H = 64 and up, M = 13–66, L = 0–12; Variable D: H = 39 and up, M = 14–38, L = 0–13.

TABLE 5  
LEVELS OF SIGNIFICANCE FOR DIFFERENCES AMONG MEANS: HIGH, MIDDLE, AND LOW GROUPS

Scale	Mothers					Fathers				
	A	B	C	D	L	A	B	C	D	L
1	.05		.03			.05				
3		.01				.05	.05			
4		.05								
6	.05	.01				.01	.01			
8	.01	.001		.01						
10	.03	.01			.05	.01				.05
11		.03	.01	.01	.01					
12	.03	.01				.001	.01	.03	.01	
14	.05	.05	.05	.05	.03		.05			
19						.03				
26									.05	.01
I	.01	.01	.05	.03		.001	.001			
II		.01				.05	.05			.01
III		.01	.05	.01	.001	.05	.05			
IV					.01					
VI	.05		.05							

*Reliability.* The reliability of each scale was estimated by the Kuder-Richardson procedure (Kuder & Richardson, 1937). These estimates are shown in Table 2. The mean reliability for mother scales and for father scales is .58. No item refinement of these scales has yet been done. The reliabilities are satisfactory for most of the scales, although a few clearly are not. One would expect the reliabilities to increase substantially once further item selection has been done. The mean reliability of the six summary scales is .71 for mothers, .74 for fathers. The total score on Models for mothers is quite unreliable, con-

sidering the number of items; this is also the summary scale with the lowest reliability for fathers.

### RESULTS

The first step in the analysis was to divide the sample into a high, middle, and low group for each of the five variables measured by the PNI. Thus each case was assigned to one of three groups on each PNI variable according to the child's score on that variable. The ranges of scores defining the high, middle, and low groups on each variable are shown in Table 4. Table 4 also shows the proportions of the cases assigned to each of the Aggression categories which fall in the respective thirds of the distributions on three other PNI variables. A similar breakdown is given for the Dependency categories. The cutting scores for assigning cases to each group were chosen so as to approximate as nearly as possible a division of the sample into equal thirds. Some consideration also was given to the distribution of scores in the total population from which this subsample was drawn.

On each of the 34 parent attitude scales, a simple analysis of variance was carried out to test the significance of the differences among the mean scores obtained by the parents of the high, middle, and low Aggression cases. This statistical procedure was deemed legitimate since the distributions of scores on the parent scales do not depart markedly from normality and variances of the three groups of cases being compared on each scale are quite similar. Separate analyses were done for mothers and for fathers. The total set of comparisons for all the attitude scales was repeated for each of the other PNI variables. Table 5 presents the results of this analysis. The  $p$  values are given for each case in which a parent attitude scale yielded a significant  $F$  for the differences in mean scores among the high, middle, and low groups.

From an inspection of Table 5, it appears that 35 comparisons of scores on mothers' scales and 22 comparisons for fathers were significant at the .05 level or better. There were 17 instances in which a parent attitude scale predicted differences in Aggression, 18 for Dependency, 7 for Withdrawal, 7 for Depression, and 8 for Likeability. The cases were originally selected, as described previ-

ously, on the basis of their scores on either Aggression or Dependency. No doubt this accounts in part for the greater number of relationships found for these two variables. It will be noted that in 16 out of 26 instances a given attitude scale is associated with more than one PNI variable. This would be expected on the basis of the intercorrelations among the PNI variables.

The relationships found between parent attitude scales and the PNI measures of deviancy (i.e., Aggression, Dependency, Withdrawal, and Depression) may be summarized as follows (the relationships with Likeability are listed separately):

Associated with all four deviancy variables:

Mothers—Low self-esteem, high ambivalence  
Fathers—High punitiveness and physical punishment

Associated with three deviancy variables:

Mothers—High consistency, high demands for aggression, high aggression and punitiveness  
Fathers—High ambivalence

Associated with Aggression and Dependency:

Mothers—High restrictiveness, high deprivation of privileges, high punitiveness and physical punishment  
Fathers—High strictness, high aggression and punitiveness, high rejection, high restrictiveness

Associated with Aggression and Withdrawal:

Mothers—High affection demonstrated, high mastery

Associated with Aggression alone:

Fathers—High affection demonstrated, high deprivation of privileges, high sex role expectations

Associated with Dependency alone:

Mothers—High rejection, high sex anxiety  
Fathers—Low self-esteem

Associated with Likeability:

Mothers—Low deprivation of privileges, low demands for aggression, low aggression and punitiveness, high self-esteem, high parental adjustment  
Fathers—Low demands for aggression, low aggression and punitiveness, high child mastery

### DISCUSSION

The major focus of this study has been on those aspects of parent attitudes found to be associated with general social deviancy in preadolescent boys. If we consider first those attitude scales related to three or all four of



the PNI measures of deviancy, the picture which emerges is one of ambivalence and punitiveness on the part of both parents. In addition, low self-esteem, high consistency, and high demands for aggression are characteristic of the mothers of the generally deviant boys. If we assume that the parent attitudes being measured by this inventory are in fact antecedent to the development of deviant behavior in the children, then we might interpret these findings as indicating that children who experience relatively intense frustration in their interactions with their parents will come to exhibit with considerable intensity a diverse set of maladaptive behaviors. These maladaptive behaviors will ordinarily include aspects of hostile aggression, overdemanding and inappropriate bids for attention, withdrawal from friendly interaction with peers, and such manifestations of sadness and distress as frequent crying.

The measure of "ambivalence," which seems to a highly significant extent to predict deviance in the children, calls for a word of explanation. It is the sum of two scales, rejection and demonstration of affection. Although these two scales correlate slightly negatively in the total population of parents, parents of children who score high on the deviance variables have, on the average, scores above the median on *both* of these scales. What meaning can be assigned to the combined score as a predictor of deviancy? Possibly the parents of the deviant children are acting in both of these somewhat contradictory ways, alternately and inconsistently being affectionate and rejecting toward their sons. It seems more likely, however, that it is primarily the rejection which is related to the deviant behavior in the children, and that these parents' elevated scores on "demonstration of affection" represent a defensive attempt at compensation for their hostile or rejecting attitudes. In such a highly educated, middle-class sample, one would be unlikely to find expressions of rejection without some concomitant manifestations of guilt and defensiveness.

In addition to punitiveness and ambivalence, the other variable most generally associated with deviance is low maternal self-esteem. This finding would seem consonant with the common view that neuroticism, mal-

adjustment, or distress in the mother is likely to prove pathogenic for the child. It seems important, however, to acknowledge the other possible interpretation of these findings, namely, that the deviations in parental attitudes, including lowered self-esteem and rejection of the child, may be the product rather than the antecedent of the child's deviant behavior. Such reactive attitudes of parents may occur where the child's pathology is primarily organic in origin. It is the basic assumption of the present study that parental attitudes may be considered as antecedent to deviant behavior in the children, but obviously neither our data nor those of any other cross-sectional study will provide us with a critical test of this hypothesis.

When we consider the attitude scales associated primarily with Aggression and Dependency, the set of parent characteristics is somewhat amplified but not essentially changed from that associated with deviance in general. It would seem that a combination of restrictiveness, punitiveness, and paternal rejection results in a pattern of hostile and demanding behavior, as reported by peers. There is some agreement with the findings of other studies as to the antecedents of Aggression and Dependency. For example, Sears, Whiting, Nowlis, and Sears (1953) found the highest level of dependency in preschool children to be associated with a combination of reward and punishment, and that aggression increased with punitiveness, up to a certain point. The findings with respect to the effect of permissiveness are contradictory. This may reflect either the difference in ages of the two samples or the more deviant manifestations of Aggression and Dependency being measured in this study.

It is of some interest that a set of father variables appears as specifically related to Aggression and not to the other PNI variables. These scales are father's sex role expectations, demonstration of affection, and use of deprivation of privileges. One might conclude that, given the general conditions for the development of deviance, if the father is interested and involved with his son, actively manipulates rewards and punishments, and is concerned that the son behave in more

masculine ways, the son's deviant behavior is more likely to take a predominantly aggressive form. In contrast, Dependency is especially related to maternal rejection and sex anxiety and low paternal self-esteem. This would suggest that more dependent forms of deviant behavior may develop in a setting in which the father is less confident and assertive and the mother reacts to the child's expression of impulse with anxious concern or aversion.

The parent characteristics associated with high Likeability, which might be considered deviance in the socially desirable direction, are in most instances the reverse of those associated with the other PNI variables. The parents of children who are especially popular and well-liked by their classmates tend not to be aggressive or punitive and have low demands for aggression. The mothers are apparently satisfied with themselves and their relationships with other family members. The fathers express high regard for their sons and evaluate them as competent.

To summarize our conclusions from these data, it would seem that deviant social behavior in boys of this age develops in a setting of punitiveness, restrictiveness, and ambivalence on the part of both parents. Within this general context of disturbed parent-child interactions, we are tempted to speculate on somewhat different roles played by mothers and fathers in the shaping of their sons' deviant behavior. The father seems to be an important influence to the extent that he is actively involved with the child and presents either an assertive and perhaps exaggeratedly masculine or a somewhat unassertive, inadequate model. The mother's impact on the son, on the other hand, seems to be a somewhat more complicated function of her internal dynamics and relationships with other family members, particularly the father. She may react toward the child with hostile rejection or anxious overconcern and restrictiveness at least in part because her own security and self-esteem are impaired. And this, in turn, may be a reflection of a disturbed or unsatisfying marital relationship. Such formulations are only very tentatively suggested by our data, including the intercorrelations among the attitude scales, but they bear some re-

semblance to Becker's (Becker, Peterson, Hellmer, Shoemaker, & Quay, 1959) conclusions, from his factor analytic study, that the attitudes of mothers and of fathers of child guidance cases were organized somewhat differently.

The sampling procedure followed in this study, as well as the properties of the PNI as a selection instrument, permitted us to study those parent characteristics associated with generally deviant social behavior within the context of the normal male elementary school population. Our findings are thus not directly comparable to studies of child guidance cases or samples of more narrowly selected deviant groups. It is perhaps worth pointing out that the degree of association among the PNI variables, as demonstrated in Table 3, suggests that "pure cases" of a particular form of deviant behavior, such as aggression or withdrawal, are relatively infrequent in the population of deviant children of this age. Perhaps, then, our findings may be considered more representative of the parental characteristics generally associated with socially deviant behavior than those found in studies of more narrowly selected groups. In any case, they are not directly comparable to the findings of studies such as Becker's (Becker et al., 1959) or Bandura's (1960) which specifically contrasted the families of aggressive with those of withdrawn children.

The question might be raised as to the need for another set of parent attitude scales when several others, particularly the PARI (Schaefer & Bell, 1958) and the Shoben (1949), are already quite widely used. Studies by Zuckerman (Zuckerman, Barrett, & Bragiel, 1960) and others (Leton, 1958) have not proved encouraging as to the usefulness of such instruments in attempts to substantiate the connections between parental characteristics and childhood psychopathology. It has seemed to us, moreover, that we might improve on available instruments. Some are too global, and they have not been specifically designed to measure those parent characteristics which have been found in previous studies of normal as well as deviant children to predict differences in specific aspects of personality development. Our choice of variables was obviously determined by a



particular theoretical point of view about personality development as well as by existing empirical evidence. The scales have the additional possible advantage that the items are worded in the natural language of parents. The findings so far would suggest that this approach will prove fruitful.

# SUMMARY

Parent attitudes were studied as they related to the social deviance of preadolescent boys. Social deviance is defined by extreme scores on the Peer Nominations Inventory, an instrument which utilizes classmates' ratings. A questionnaire composed of 28 rationally constructed scales referring to aspects of family relationships and child rearing practices was administered to samples of mothers and fathers.

The five social reputation scales of the PNI are Aggression, Dependency, Withdrawal, Depression, and Likeability. The first four are negatively related to Likeability and intercorrelate positively to define a cluster of generally maladaptive social behaviors. Parent attitudes which differentiated deviant from nondeviant boys included ambivalence, punitiveness, demands for aggression, restrictiveness, and low maternal self-esteem. High Likeability was in general associated with the opposite parent characteristics. In addition, the mothers of the popular boys reported high parent adjustment and the fathers gave more favorable evaluations of their sons' competence.

Additional findings suggested that fathers play an important role in the development of deviant behavior in general, but contribute uniquely in several ways to aggression. Fathers of more aggressive boys reported more

affection and manipulation of rewards and more stereotyped sex role expectations. A different set of parent characteristics seems specifically related to dependency, namely, mothers' rejection and sex anxiety and fathers' low self-esteem.

# REFERENCES

- BANDURA, A. Relation of family patterns to child behavior disorders. Progress Report, June 15, 1960, USPHS Grant M-1734.
- BECKER, W. C., PETERSON, D. R., HELLMER, L. A., SHOEMAKER, D. J., & QUAY, H. C. Factors in parental behavior and personality as related to problem behavior in children. *J. consult. Psychol.*, 1959, 23, 107-118.
- KUDER, G. F., & RICHARDSON, M. W. The theory of estimation of test reliability. *Psychometrika*, 1937, 2, 151-160.
- LETON, D. A. A study of the validity of parent attitude measurement. *Child Developm.*, 1958, 29, 515-520.
- MILLER, D. R., & SWANSON, G. E. *Inner conflict and defense*. New York: Holt, 1960.
- SCHAEFER, E. S., & BELL, R. Q. Development of a Parental Attitude Research Instrument. *Child Developm.*, 1958, 29, 339-362.
- SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child-rearing*. Evanston, Ill.: Row, Peterson, 1957.
- SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. psychol. Monogr.*, 1953, 47, 135-234.
- SHOBEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. psychol. Monogr.*, 1949, 39, 101-148.
- WINDER, C. L., & WIGGINS, J. S. Measurement of social adjustment in preadolescent boys. Progress Report, November 7, 1960, USPHS Grant M-2745 (C1).
- ZUCKERMAN, M., BARRETT, B. H., & BRACIEL, R. M. The parental attitudes of parents in child guidance cases: I. Comparisons with normals, investigations of socioeconomic and family constellation factors, and relations to parents' reactions to the clinics. *Child Developm.*, 1960, 31, 401-417.

(Received April 5, 1961)



# PREFERENCE FOR DELAYED REWARD AS A FUNCTION OF AGE, INTELLIGENCE, AND LENGTH OF DELAY INTERVAL<sup>1</sup>

WALTER MISCHEL  
Stanford University

AND

RALPH METZNER  
Harvard University

The ability to delay immediate gratification has long occupied a central role in theorizing about child development (e.g., Freud, 1959; Rapaport, 1950; Singer, 1955). Only recently, however, has this type of behavior been subjected to experimental research. One group of studies (e.g., Singer, Wilensky, & McGraven, 1956), using primarily adult schizophrenic subjects, has explored the correlates of "delaying capacity" as reflected in the Rorschach human movement (*M*) response. A second group of studies has approached the problem of delay of gratification directly by eliciting a choice between an immediate, smaller reward (*ImR*) and a delayed, larger one (*DelR*). Thus far, using children from other cultures as subjects, preference for *DelR* has been related positively to social responsibility (Mischel, 1961c), to need for achievement and "naysaying" (Mischel, 1961a), and to the presence of the father in the home under certain conditions (Mischel, 1961b).

The present study attempts to isolate further some of the basic parameters of the preference for delayed as opposed to immediate reward in a simple choice situation. Specifically, three main hypotheses were tested, viz., that preference for *DelR* should be related (a) positively to age and (b) to intelligence, and (c) negatively to length of the delay interval.

The first hypothesis is derived from the psychoanalytic theory of development, according to which the transition from immediate wish fulfilling types of behavior to delayable reality testing types is the crucial element in the change-over from primary to

secondary processes. It is assumed that this transition is not an all-or-none affair, but a continuing process resulting in behavior which can be ranged on a continuum from uncontrollable impulsivity to compulsive postponement. The ability to delay gratification in various need areas is presumed to increase with age, and hence in a simple, paradigmatic choice situation older children should more often prefer the delayed, larger as opposed to the immediate, smaller reward. The second hypothesis is derived from the same source, more specifically from the psychoanalytic theory of thinking (Rapaport, 1950), according to which learning to delay is intimately bound up with learning to think; i.e., the use of cognitive reality testing replaces the use of uncontrolled motor discharge in the attainment of needs. Hence we would expect greater delaying capacity to be reflected in greater cognitive facility, i.e., intelligence. The third hypothesis is a deduction from any expectancy theory which assumes (a) that the effective value of an incentive is a joint function of its "real" value and the expectation of getting it, and (b) that the expectation of getting something decreases with its temporal distance:

Previous empirical studies, using Trinidadian children (Mischel, 1958) and Palestinian Arab refugee children (Melikian, 1959), have found evidence for a relationship between *DelR* preference and age, using age ranges from 7 to 9 and from 5 to 10, respectively. The present study extends the age range by using a single measure on children from 5 to 12 years old. Melikian also found a suggestive relationship with intelligence (using Goodenough's Draw-A-Man Test); and Levine, Spivack, and Wight (1959) have summarized a large body of data correlating the Rorschach *M* response, their indirect measure of delaying capacity, with intelligence in

<sup>1</sup> This study was supported in part by a PHS research grant (M 3344) from the National Institute of Mental Health, United States Public Health Service, and a grant from the Laboratory of Social Relations, Harvard University. Thanks are also due to Mark Harris for help in data collecting.

adults. It should be noted however that the relationship of *M* to the ability to delay motor discharge has not so far been confirmed for adolescents or children (Spivack, Levine, Fuschillo, & Tavernier, 1959). No previous work has apparently been done on the third hypothesis.

In addition, the present experiment seeks to clarify the relationship of delaying capacity to length of future time perspective. Previous work, using adolescent subjects, and a measure of "cognitive inhibition," has found that subjects high on inhibitory ability have longer future time "conceptions" (Levine, Spivack, Fuschillo, & Tavernier, 1959). But Mischel (1961c), using the direct measure employed in the present study and Trinidadian children, found that preference for DelR was related not to length, but to *accuracy* of time conceptions (of past events). Theoretically, there are two different things one could expect. If we assume that "immediate" subjects are unable to tolerate long time delays, we would expect a linear relationship between length of time perspective and delay preference. Realistically, however, temporal distance can be overestimated as well as underestimated. Therefore, assuming that "delay" subjects have greater reality contact, we would expect a curvilinear relationship between length of time perspective and delay preference, with "immediate" subjects making estimates that are either very short or very long, in terms of plausible realism.

## METHOD

### Subjects

Subjects were 162 elementary school children, from predominantly lower-middle and lower-class urban backgrounds in the Boston area. The 126 subjects (68 boys, 58 girls) of the main group (Grades 1 through 6) were from one school and were all tested on one day. In addition, 36 subjects (21 boys and 15 girls, all kindergarten children from a second, similar school in the same community), were tested separately on another day.

### Main Group

**Design.** The purpose of the experiment called for a  $6 \times 5$  design, in which age was varied by using six grades and length of delay interval was varied by using five different delay periods (1 day, 5 days, 1 week, 2 weeks, 4 weeks). On the basis of the school list, each class was randomly divided into

five groups and each group assigned to one of the delay periods. Due to absenteeism and to the fact that the younger grades had fewer children, some of the cells had very small frequencies. Subjects from Grades 3 through 6 and Grades 1 through 2 were tested together, though subjects from each delay length category were tested separately. Thus for the purpose of administering the procedure there were 10 groups and these were assigned at random to two experimenters (*Es*). However, since the groups made up of the older children (Grades 3 through 6) were larger, one *E* tested more subjects than the other (74 vs. 52). The possible effects of this will be discussed later.

**Procedure and measures.** The reinforcements, chosen on the basis of pretesting and previous experience, were: a \$.05 candy bar (Hershey, with almonds) today (ImR) or a \$.10 candy bar of the same kind later (DelR, e.g., tomorrow or next week, etc.). This real choice was offered, as a reward for cooperating, at the end of the total procedure. Both candies were shown and at the promised time the subject was given the one he actually chose.

In order to obtain a measure of *future time perspective* the following questions were asked:

1. What year do you think the first man will reach the moon?
2. What year do you think somebody will run the mile in 3.5 minutes?
3. What year will everyone have their own airplanes, like today they have their own cars?
4. What year will somebody hit 60 home runs again?
5. How old will you be when you are grown up?
6. How old will you be when you get married?
7. How old will you be when you make the most money?
8. When will the summer vacation start?

Since their distributions showed very little spread, Questions 5 and 8 were omitted from the analysis. The answers to the other six questions were divided into thirds and subjects assigned scores from 1 (shortest) to 3 (longest) on each question; these were then summed to get a total length of time perspective score.<sup>2</sup>

### Additional Group

**Procedure.** The kindergarten subjects were included in order to extend the age range downward

<sup>2</sup> For the younger children (Grades 1 and 2) only Questions 1 and 6 were included in the procedure, as a first step in pretesting appropriate items for this age range. So the analysis of time perspective is based on the older subjects only. Brief simple measures of acquiescence and occupational preference were also included in the procedure for the older subjects. However, since these were exploratory and are not relevant to the main purpose of the study, they will not be discussed in the present report.



as far as possible. The only procedure here was to ask subjects to print their first name on a piece of paper. They were told to indicate their choice of reward—\$.05 candy today, or \$.10 candy next week) by writing T (today) or L (later) on their paper, the writing being first demonstrated on the blackboard. Subjects were tested in two groups by one experimenter ( $E_1$ ).

### RESULTS

The data were examined for sex differences, which were not significant ( $\chi^2 = 2.306$ ,  $df = 1$ ,  $p > .10$ ); therefore all subsequent analyses combine results from boys and girls.

The results on the two main dimensions, age and length of delay interval, are shown in Table 1. There is a clearly visible trend for the proportion of DelR choices to increase with age, the most marked change occurring between Grades 3 and 4. The results from the kindergarten subjects (mean age = 5.7) are consistent with those of the main group: the ratio of immediate choices to delayed is 26:10. If we consider only the age variable we have a set of proportions in seven ordered categories (Grades 1 through 6 and kindergarten). Hence a statistic that enables us to take out for trend is appropriate (Armitage, 1955). The chi square for linear regression is 23.75 ( $df = 1$ ,  $p < .001$ ). The nonlinear effect is not significant ( $\chi^2 = 9.04$ ,  $df = 5$ ,  $p > .10$ ). Alternatively, we may do a correlational analysis, in which chronological age is one variable and choice of DelR the other. This yields a point biserial  $r$  of .65 ( $N = 160$ ,  $p < .0001$ ). The 95% confidence limits for this correlation coefficient are .55 and .73.

The mean IQ of children choosing DelR is 105.7 ( $N = 64$ ,  $SD = 11.25$ ), and of children choosing ImR 99.0 ( $N = 58$ ,  $SD = 10.78$ ).<sup>3</sup> This difference is highly significant ( $CR = 3.37$ ,  $p < .0005$ ). Again, we may perform an alternative analysis in terms of correlation: the point biserial  $r$  between IQ and DelR choice is .29 ( $N = 122$ ,  $p < .0005$ ). The 95% confidence limits for this correlation coefficient are .12 and .45. In order to test whether the correlation between preference

TABLE 1

THE RELATIONSHIP OF AGE AND DELAY INTERVAL LENGTH TO PREFERENCE FOR DELAYED VS. IMMEDIATE REWARD<sup>a</sup>

Delay interval	Grades <sup>b</sup>						Totals <sup>a</sup>
	1 (6.6)	2 (7.6)	3 (8.8)	4 (9.5)	5 (10.6)	6 (11.6)	
1 day	3:0	2:3	1:1	0:4	0:5	0:5	6:18
5 days	4:1	1:1	4:0	1:4	2:3	2:3	14:12
1 week	2:2	2:1	4:1	1:4	2:3	2:3	13:14
2 weeks	3:1	3:1	2:2	1:3	0:3	0:5	9:15
4 weeks	2:0	2:2	4:0	4:2	3:1	2:3	17:8
Totals <sup>d</sup>	14:4	10:8	15:4	7:17	7:15	6:19	59:67

<sup>a</sup> In each cell the first figure gives the number of immediate, smaller (ImR) choices, the second the number of delayed, larger (DelR) choices.

<sup>b</sup> The figures in parentheses are mean chronological ages.

<sup>c</sup>  $\chi^2$  for linear trend across delay intervals = 4.91,  $df = 1$ ,  $p < .05$ .

<sup>d</sup>  $\chi^2$  for linear trend across age groups = 23.75,  $df = 1$ ,  $p < .001$ .

for DelR and intelligence may be explained in terms of the correlation between preference for DelR and age (or vice versa) the correlation between age and IQ for this sample was computed. Since this was not significant ( $r = .14$ ,  $N = 122$ ,  $.10 > p > .05$ ), it may be assumed that the two correlations are independent.

To test the third hypothesis the total proportions in the five delay length categories (last column of Table 1) were analyzed in the same way as the proportions in the age categories. The chi square for linear regression is 4.91 ( $df = 1$ ,  $p < .05$ ). The nonlinear effect is again not significant ( $\chi^2 = 5.35$ ,  $df = 3$ ,  $p > .10$ ). Thus the hypothesis is confirmed: the proportion of ImR choices increases linearly with length of delay interval.

In order to test for interaction between age and length of delay interval, the analysis of trend due to delay length was also carried out separately for younger and older children. If we consider only the subjects from the first to the third grade (i.e., the totals of the first three columns in Table 1) the chi square for linear regression is 0.23 and that for curvilinearity is 1.69, neither of which is significant. In the subjects drawn from the fourth to the sixth grade (i.e., the totals of the last three columns of Table 1) the chi square for linear regression is very significant ( $\chi^2 = 7.08$ ,  $df = 1$ ,  $p < .01$ ) and the nonlinear ef-

<sup>3</sup> The IQ scores are based on the Lorge-Thorndike Form A for the first and second grade subjects, and on the Pintner General Ability Test, Verbal Series, Forms A or B, for the older subjects. All scores were taken from the school records.



TABLE 2

THE RELATIONSHIP OF FUTURE TIME PERSPECTIVE TO PREFERENCE FOR DELAYED REWARD

Condition	Time perspective <sup>a</sup>			
	Short	Medium	Long	
ImR	16	5	9	$\chi^2=11.72, p<.01$
DelR	9	20	22	

<sup>a</sup> Categorization of the distribution into closest approximation of thirds.

fect is slightly significant ( $\chi^2 = 8.66, df = 3, p < .05$ ). Thus it would appear that the overall trend in all groups combined is attributable to the older group only, and that the younger children make no differentiation between the various time periods. However, the interpretation of this interaction is not unambiguous in view of the experimenter differences to be discussed below.

The data on length of time perspective into the future, using Grades 3 through 6 only, yield a mean of 11.27 ( $SD = 2.85$ ) for the group choosing the immediate, smaller reward ( $N = 30$ ); and a mean of 12.20 ( $SD = 1.92$ ) for the group choosing the delayed larger reward ( $N = 51$ ). Thus time perspective is slightly, though nonsignificantly, shorter in the ImR group ( $t = 1.57, p > .10$ ), but significantly more variable ( $F = 2.21, p < .02$ ).

In order to examine more closely the nature of the difference in variability, the distribution of the total time perspective scores was divided into thirds and thus compared with ImR vs. DelR choice. This is shown in Table 2, where it can be seen that there is a significant tendency for the ImR subjects to have either very low or very high time scores. In terms of individual questions this means, e.g., that they were more likely to give answers such as "this July" or "in the year 30,000." There are no significant differences

TABLE 3

PREFERENCE FOR DELAYED VS. IMMEDIATE REWARD IN SUBJECTS TESTED BY THE TWO EXPERIMENTERS

Condition	$E_1$	$E_2$	
ImR	17	42	$\chi^2 = 6.0, p < .02$
DelR	35	32	

TABLE 4

RELATIONSHIP OF DELAY PREFERENCE TO AGE, FOR EACH  $E$  SEPARATELY

Condition	$E_1$			$E_2$		
	Grades			Grades		
	1 and 2	5 and 6		1 and 2	5 and 6	
ImR	13	0	$\chi^2=15.01,$ $p<.001$	11	13	$\chi^2=4.41,$ $p<.02$
DelR	7	18		5	26	

in the time perspective scores for the different age groups. The correlation of time perspective length with IQ is .17 ( $N = 76, p < .07$ ) which is suggestive but not strong.

### Differences between $E$ s

Table 3 shows that there is a significant difference between  $E$ s:  $E_1$  elicits more DelR choices,  $E_2$  elicits more ImR choices. If both experimenters tested the same numbers and kinds of subjects this difference would not affect the interpretation of the results. However, this was not possible both because of sample shrinkage and because the procedure called for first and second grades to be tested separately from the older children. Specifically, each  $E$  tested five groups, one at each delay interval— $E_1$  tested three younger groups and two older,  $E_2$  tested three older groups and two younger. In addition, the first and second grades were smaller than the others. The result was that  $E_2$  tested more older children than  $E_1$ , and also more children at the longer delay periods. In order to determine to what extent differences in  $E$ s affected the previously found relationships,

TABLE 5

RELATIONSHIP OF DELAY PREFERENCE TO DELAY LENGTH, FOR EACH  $E$  SEPARATELY

Condition	$E_1$			$E_2$		
	1 day	4 weeks		1 day	4 weeks	
ImR	1	4	$\chi^2 = 5.96, p < .02$	5	13	$\chi^2 = 0.02, ns$
DelR	15	2		3	6	

these were reanalyzed for each  $E$  separately.

Table 4 shows that if preference for DelR in the two youngest and two oldest age groups is compared, the results are significant for both  $E$ s. Moreover it should be noted that the difference in experimenters would go against the hypothesis, i.e., reduce the strength of the relationship between age and preference for DelR, since  $E_2$ , who elicited more ImR choices, also tested more subjects in the older age groups.

Table 5 shows that if preference for DelR is compared in the groups tested under the shortest (1 day) and longest (4 weeks) delay periods, the difference is significant for  $E_1$  but not for  $E_2$ . It may be concluded that the tendency of  $E_2$  to elicit more ImR choices masked the effect of delay length differences in subjects tested by him, and that the overall trend must be regarded as a joint function of differences in length of delay interval and reaction to the  $E$ .

Finally, the previously discussed interaction between age and delay interval is absent for subjects tested by  $E_1$ , and nonsignificant although in the predicted direction for subjects tested by  $E_2$ . On the basis of the present data then, the previous conclusion that for the younger children only the "now-later" contrast is relevant must be regarded as tentative. The interaction between age and delay length appears in the data, but is not unambiguously interpretable because it is confounded with  $E$  differences. We may state the conclusion to be drawn as follows:

1. There is a general tendency for the proportion of DelR choices to decrease with increasing length of the delay interval.
2. This tendency is especially marked in the older subjects, but this may be due to either of two possibilities: (a) older subjects make more sensitive temporal discriminations, or (b) a general set to make more ImR choices to the experimenter ( $E_2$ ) who tested most of the older subjects.

#### DISCUSSION

As predicted, a strong linear relationship of preference for DelR with age, and a less strong, but still very significant relationship with intelligence, were found. Furthermore,

the predicted inverse relationship of delay preference and delay interval was also evident, although it was at least partially masked by differences in the reactions to the two experimenters. It should be noted that the latter finding is in accord with those of Mahrer (1956) and Terrell (1959) who found that the experimental manipulation of incentive values in young children is extremely sensitive to differences in the behavior of the  $E$ .

The inverse relationship of delay preference and length of delay interval was predicted on the basis of the assumption that the expectancy of obtaining a reinforcement decreases with its temporal distance. However, the present experiment does not rule out the alternative explanation that the negative value of waiting subtracts from the positive value of the incentive, regardless of expectancy. The experiment by Mahrer (1956) showed that different levels of expectancy affect preference for delayed rewards even when delay length is held constant. The converse experiment, varying the length of the delay interval while holding expectancy constant, has yet to be done. The results of the present experiment may be regarded as a preliminary control for such a study: they show that length of delay interval affects DelR preference. But this may reflect differences in expectancy only, if the passage of time per se is unimportant, as Rotter (1954, p. 161) has suggested; or it may reflect differences in expectancy and differences in the aversiveness of waiting.

In the present experiment the major changes occurred at the age of 8.5-9, between the third and fourth grades. The change from a preponderance of ImR choices to a preponderance of DelR choices came at this point, and there was some suggestive evidence that the effect of delay intervals of differing length was significant above this age only. It is possible to speculate that for younger children the different lengths of time are meaningless and only the now-later contrast is relevant. The present data lend some support to such a notion, though it is doubtful whether the precise age at which the change occurs is a general "critical age." For, in the first place, different pairs of reinforcements would probably yield different cutting-points, even



though the overall trend might remain the same. Melikian (1959), for example, found that using his particular monetary reinforcements (10 fils vs. 20 fils) in his group of refugee children, the crucial age was 6, and concluded that after 6, no more increase in preference for delay is to be expected, since refugees eventually lose hope. In the present study, change occurred after 6, but none after 9. It would seem more plausible to interpret such results as reflecting the fact that the differences in the particular reinforcements chosen cease to be very important after a certain age. Naturally, this age will vary with the values chosen. One might expect that the "critical age" could be located at any desired point on the age scale by appropriate manipulation of reinforcement values. The problem in demonstrating an extended age trend is one of finding the appropriate *measure*, rather than one of locating specific stages in development. Secondly, the time perspective scores show no significant change at 8 or 9, as one would expect if this were indeed the general critical age.

The results on time perspective tend to support the second of the two possibilities mentioned earlier, viz., that delay subjects make moderate, realistic estimates of future events, whereas immediate subjects make either extremely short or extremely long estimates. There is also a slight tendency for the immediate subjects to have shorter time perspective. Also, the tendency for the more intelligent subjects to have longer time perspectives confirms the findings of Levine, Spivack, Fuschillo, and Tavernier (1959) on adolescents. That these results may be further complicated by a "ceiling effect" in the delay measure is suggested by some recent findings (Mischel, Gershuni, & Vollmar, 1960). Using a continuous rather than a dichotomous delay preference measure and a measure of length of time span of TAT stories, it was found that subjects with the highest and lowest DelR scores, i.e., those located at the extremes of the distribution of DelR scores, have shorter time spans than those with medium DelR scores. This effect would blur differences in time perspective (assuming the two time measures are roughly analogous) when a crude measure of delay preference

such as that of the present study is used. For in the DelR group there would be included both subjects with long time scores (the moderate delayers) and subjects with short time scores (the extreme delayers).

## SUMMARY

Using a simple choice measure of preference for delayed reward on elementary school children in the Boston area, ranging in age from 5 to 12, and under five different lengths of delay interval, the following hypotheses were tested and confirmed: (a) preference for delayed reward is positively related to age; (b) positively to intelligence; (c) and negatively to length of the delay interval. The findings on Hypothesis c were not entirely unambiguous owing to the extreme sensitivity of the delay measure to differences in the *E*. An additional finding was that subjects preferring the immediate reward tend to have more variable future time perspectives and that length of time perspective is slightly related to IQ, but not to age.

## REFERENCES

- ARMITAGE, P. Tests for linear trends in proportions and frequencies. *Biometrics*, 1955, **11**, 375-386.
- FREUD, S. Formulations regarding the two principles in mental functioning. (Originally published 1911) In, *Collected papers*. Vol. IV. New York: Basic Books, 1959. Pp. 13-21.
- LEVINE, M., SPIVACK, G., FUSCHILLO, JEAN, & TAVERNIER, ANN. Intelligence, and measures of inhibition and time-sense. *J. clin. Psychol.*, 1959, **15**, 224-226.
- LEVINE, M., SPIVACK, G., & WIGHT, B. The inhibition process, Rorschach human movement responses, and intelligence: Some further data. *J. consult. Psychol.*, 1959, **23**, 306-313.
- MAHRER, A. The role of expectancy in delayed reinforcement. *J. exp. Psychol.*, 1956, **52**, 101-105.
- MELIKIAN, L. Preference for delayed reinforcement: An experimental study among Palestinian Arab refugee children. *J. soc. Psychol.*, 1959, **50**, 81-86.
- MISCHEL, W. Preference for delayed reinforcement: An experimental study of a cultural observation. *J. abnorm. soc. Psychol.*, 1958, **56**, 57-61.
- MISCHEL, W. Delay of gratification, need for achievement, and acquiescence in another culture. *J. abnorm. soc. Psychol.*, 1961, **62**, 543-552. (a)
- MISCHEL, W. Father-absence and delay of gratification: Cross-cultural comparisons. *J. abnorm. soc. Psychol.*, 1961, **63**, 116-124. (b)



- MISCHEL, W. Preference for delayed reinforcement and social responsibility. *J. abnorm. soc. Psychol.*, 1961, **62**, 1-70. (c)
- MISCHEL, W., GERSHUNI, C., & VOLLMAR, J. Studies on delay of gratification and time sense. Unpublished manuscript, Harvard University, 1960.
- RAPAPORT, D. On the psychoanalytic theory of thinking. *Int. J. Psychoanal.*, 1950, **31**, 161-170.
- ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
- SINGER, J. L. Delayed gratification and ego-development: Implications for clinical and experimental research. *J. consult. Psychol.*, 1955, **19**, 259-266.
- SINGER, J. L., WILENSKY, H., & MCGRAVEN, VIVIAN G. Delaying capacity, fantasy, and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, **20**, 375-383.
- SPIVACK, G., LEVINE, M., FUSCHILLO, JEAN, & TAVERNIER, ANN. Rorschach movement responses and inhibition processes in adolescents. *J. proj. Tech.*, 1959, **23**, 462-466.
- TERRELL, G. Manipulatory motivation in children. *J. comp. physiol. Psychol.*, 1959, **52**, 705-709.

(Received April 6, 1961)

## ANXIETY AND AVOIDANCE BEHAVIOR IN SCHIZOPHRENICS IN RESPONSE TO PARENTAL FIGURES<sup>1</sup>

JAMES C. BAXTER<sup>2</sup> AND JOSEPH BECKER<sup>3</sup>

*John Umstead Hospital, Butler, North Carolina*

The cardinal importance of deviant familial relationships in the background of schizophrenic patients has long been recognized by personality theorists. A sizable literature in the area strongly suggests that the schizophrenic comes from a family milieu differing in systematic ways from that of his non-schizophrenic peer (Bateson, Jackson, Haley, & Weakland, 1956; Bowen, 1959; Lidz & Fleck, 1960; Wynne, Ryckoff, Day, & Hirsch, 1958). Recent reports have also drawn attention to differences in patterns of family relationships which can be distinguished between subgroups of schizophrenics. By differentiating patients along a continuum of premorbid personality adequacy, as defined by the Phillips (1953) Scale of Premorbid Adjustment, Rodnick and Garmezy (1957; Garmezy & Rodnick, 1959) and their associates have identified reliable patterns of asymmetrical authority and submission, as well as differences in interpersonal conflict in the family. According to these studies, male schizophrenics who have attained a relatively high degree of psychosocial maturity and sexual participation prior to their psychosis (Goods) tend to come from homes dominated by a tyrannical father, with the mother occupying a comparatively weak and ineffectual role. On the other hand, patients who have failed to establish stable sexual relationships and who have been relative social isolates prior to the appearance of the disorder (Poors) tend to come from mother dominated homes high in conflict and interpersonal aggression (Farina, 1960; Garmezy, Farina, & Rodnick, 1960; Garmezy & Rodnick, 1959; Garmezy, Stockner, & Clarke, 1959; Rodnick & Garmezy, 1957). The premorbid adequacy variable has also been found to bear a significant relation-

ship to differences in reaction sensitivity to social censure, which, in turn, has been shown to relate to performance deficits on a number of experimental variables. This line of investigation has suggested that a highly generalized avoidance response is more prepotent in Poors than in Goods, as a function of increased sensitivity to censure, and that this response tendency functions as an interference to integrated problem solving behavior. Rodnick and Garmezy (1957) have proposed a relationship between the differences in familial backgrounds in these groups and their differential sensitivity to censure and rebuff (p. 146).

The present study was aimed at evaluating several implications of this position in terms of anxiety and avoidance behavior aroused in schizophrenics in response to parental figures. If exposure to differing patterns of parental dominance and conflict has a significant effect in producing the reaction sensitization determining the behavior deficits observed, then differences in anxiety and avoidance should emerge in thematic productions developed in response to pictorial figures symbolizing these early traumatic (censorious) interactions. Accordingly, the following predictions were made: (a) Poor premorbid schizophrenics should produce more anxiety related thematic imagery in response to the maternal figure than to the paternal figure, while Goods should show the opposite effect; (b) Poors should produce a significantly greater amount of avoidance related imagery in response to both parental figures; and (c) Poors should produce more avoidance in response to the mother than the father, with Goods showing the reverse pattern.

### METHOD

*Subjects.* The sample was chosen from a pool of 66 male patients undergoing treatment at the Veterans Administration Hospital, Roanoke, between 1949 and 1952, who had been given the diagnosis of schizophrenia by qualified staff psychiatrists, and

<sup>1</sup> This research was supported in part by National Institute of Mental Health Grant OM-124, "Family Centered Treatment in Schizophrenia."

<sup>2</sup> Now at the University of Kentucky.

<sup>3</sup> Now at the University of Illinois.

who had given responses to cards 6BM (mother-son) and 7BM (father-son) of the TAT during routine, individually administered psychological testing.<sup>4</sup> The length of time in the hospital before testing was less than 30 days in all cases. Only subjects between 20 and 40, who had been reared by an intact family until the age of 18 were included.<sup>5</sup> Patients whose diagnosis was complicated by indications of mental deficiency or organic brain pathology or who had received radical somatic therapy within 90 days before testing or were currently undergoing regular drug therapy were excluded. Some patients occasionally taking small amounts of sedatives to facilitate sleep were included, if the most recent administration was no later than the night before testing. The number of such patients included was small, and no difference in the proportion between groups was noted.

Thirty schizophrenics fulfilled all criteria imposed. Premorbid personality ratings were obtained from the case history of each patient by rating the degree and adequacy of social participation and sexual behavior prior to the onset of psychosis on the Phillips (1953) Scale of Premorbid Adjustment. This scale differentiates patients along a 30-point continuum of premorbid adequacy, with higher scores reflecting greater degrees of social isolation and sexual disturbance. The experimental groups were organized by taking contrasting extremes on this dimension. Patients with a score of 15 or less were assigned to the group of Goods ( $N=12$ ), while patients obtaining scores greater than 15 were classified as Pours ( $N=18$ ). The ranges of the ratings obtained were 3-15 and 17-27 for Goods and Pours, respectively. Mean age and years of education for both groups are presented in Table 1. Inspection of these data reveals that the groups are quite comparable on these variables, with no significant differences appearing. Less precise data also suggested comparability between groups for diagnostic subclassification of schizophrenia and socioeconomic position.

**Scoring procedure.** Only responses to Cards 6BM and 7BM of the TAT were considered in the analysis. These cards were chosen for study because they typically elicit thematic imagery relevant to parent-child relationships, and they are most frequently interpreted directly as mother-son and father-son interactions, respectively. A rating schedule was devised which would reflect various component dimensions

<sup>4</sup> We are indebted to Lon Ussery, University of North Carolina, for making the protocols and necessary information available to us and for doing the ratings of premorbid adjustment. Burk Smith, Veterans Administration Hospital, Roanoke, Virginia, also provided invaluable assistance in obtaining drug and somatic therapy information on the patients.

<sup>5</sup> An intact family was defined by the absence of social history information indicating death, separation, or divorce of the parents before the patient had reached age 18. Fifty percent of the Goods and 18% of the Pours who were otherwise eligible were discarded on the basis of this criterion.

TABLE 1

MEAN AGE AND EDUCATIONAL LEVEL OF GOOD AND POOR PREMORBID GROUPS

Level	N	SD	M
Age			
Goods	12	4.89	27.9
Pours	18	4.26	27.4
Education			
Goods	12	2.86	10.2
Pours	18	2.79	10.4

of anxiety and avoidance behavior, and would permit reliable judgments to be made from oftentimes brief, elliptically recorded TAT protocols. Both stylistic and contentual aspects of the protocols were considered relevant to the variables under study, and the weightings given the several dimensions chosen roughly corresponded to the experimental and clinical validity attributed to them.

A number of style and content variables in the linguistic productions of a variety of populations have been shown to be significantly related to manifest anxiety (e.g., Dollard & Mowrer, 1947; Mahl, 1959a; Meadow, Greenblatt, & Solomon, 1953; Osgood, 1957; Phillips & Smith, 1953; Walker & Atkinson, 1958). Five areas which seem to have achieved most extensive validation and clinical utility were chosen for inclusion in an anxiety rating schedule. The areas chosen were: indications of verbal interferences, viz., repetitions, sentence incompletions, sentence corrections, tongue slips, and stuttering, as defined by Mahl (1959b); statements of inability to comply with the task; signs of disorganization of the theme, e.g., digressions, irrelevancies, "private" associative connections, etc.; content of a predominantly negative affective tone; and the development of a bleak, pessimistic outcome. Variable weights assigned to components of each of the areas were combined into a 12-point anxiety index, with higher scores indicating greater anxiety. A comparable index of avoidance related behavior was developed by translating the results of several divergent studies into scales applicable to TAT protocols (e.g., Bindra, 1959; Miller, 1959; Scott, 1958). Four dimensions were chosen and assigned varying weights. The general categories were: the amount of displacement of person, place, and/or time evident in the theme; the predominant direction of movement implied, i.e., toward or away from an affectively meaningful, spatially proximal relationship; the amount of verbal productivity in thought units (Dollard & Mowrer, 1947); and the degree to which objectively present figures were ignored in the themes produced. A 12-point avoidance index was obtained by summing component scores of the several categories, with higher values indicating a greater degree of avoidance.

All ratings were made by one judge from transcripts of the original protocols, coded to prevent identification of group membership. Two additional



TABLE 2  
ANALYSIS OF THE VARIANCE OF ANXIETY  
INDEX SCORES

Source	SS	df	MS	F
Between-patients	166.48	29		
Premorbid level	2.84	1	2.84	<1.00
Error	163.64	28	5.84	
Within-patients	156.50	30		
Parental figure	12.15	1	12.15	2.64
Premorbid $\times$ Parent	15.21	1	15.21	3.30**
Error	129.14	28	4.61	
Total	322.98	59		

\*\* Significant at the .05 point, one-tailed.

raters also independently scored 15 randomly chosen protocols in order to estimate interrater reliability.<sup>6</sup> Product-moment correlations between raters were found to vary from .70 to .88 for the anxiety index, and from .93 to .97 for the avoidance index, with mean correlations of .78 and .95, respectively. The reliability of the anxiety scores was somewhat lower than anticipated, although it was considered adequate for present purposes, in view of the protean nature of content expressions of anxiety and the variability in quality of recording found in the TATs. The reliability of the avoidance measure, on the other hand, was quite adequate.

### RESULTS

The first hypothesis predicted that the pattern of anxiety related imagery would differ within the responses of Goods and Poors, with Poors expressing more anxiety in response to the mother figure and Goods more in response to the father figure. Accordingly, the anxiety scores for both groups were entered in a Type I mixed analysis of variance table (Lindquist, 1953). The assumption of homogeneity of score variance within cells was evaluated by Hartley's  $F_{\max}$  test (Edwards, 1954) and found to be tenable ( $p > .05$ ). A summary of the analysis of anxiety scores is presented in Table 2. It can be seen that the predicted interaction of anxiety reactivity and parental figure is significant at the .05 point of confidence.<sup>7</sup> It is also apparent that the over-all level of anxiety between groups and mean responsiveness to the mother as opposed to the father figure do not differ significantly. The

<sup>6</sup> We are indebted to Walter Hooks, John Umstead Hospital, for assisting the authors in making the ratings.

<sup>7</sup> One-tailed tests of significance are reported for the hypotheses, since directional predictions were made in each case.

mean anxiety scores for the groups are represented graphically in Figure 1. The curves of anxiety responsivity interact as predicted, with Poors reflecting greater responsivity to the mother than the father figure, while the pattern for Goods is in the opposite direction.

The avoidance scores permitted evaluation of the remaining hypotheses. That is, the over-all level of avoidance was expected to be greater for Poors, and the patterns of avoidance behavior were expected to differ between Goods and Poors. Scores were arrayed in a Type I table, and tested for homogeneity of cell variances. Hartley's  $F_{\max}$  test indicated that such an assumption was questionable ( $.10 > p > .05$ ), so that a square root transformation of all scores was done in order to remove heterogeneity within cells. These transformed scores were then analyzed for group and pattern differences. Table 3 summarizes the results of these comparisons. It can be seen that a significant difference in the level of avoidance behaviors exists between groups ( $p < .025$ ), with Poors exceeding Goods in response to both figures. The expected difference in the patterning of avoidance behavior did not occur, however ( $p > .05$ ). Inspection of individual means indicated that, while Goods tended to show the expected increase in avoidance responsivity to the father figure, Poors were equally avoidant in response to both parental figures.

It is possible that the failure of discriminative avoidance could be accounted for as a result of the excessively high level of avoidance behavior in the Poors. It could be ar-

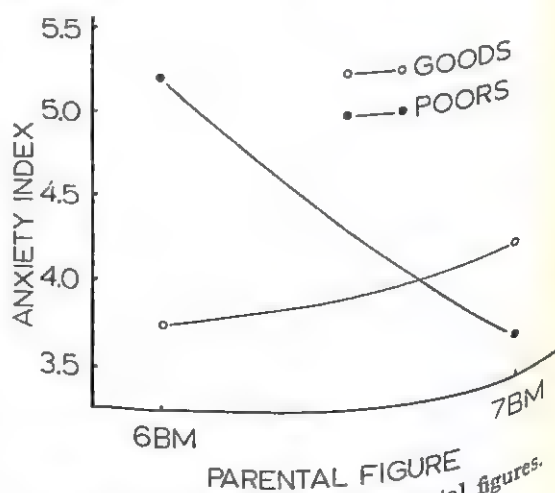


FIG. 1. Anxiety responsivity to parental figures.

TABLE 3

ANALYSIS OF THE VARIANCE OF AVOIDANCE INDEX SCORES

Source	SS	df	MS	F
Between-patients	7.78	29		
Premorbid level	1.07	1	1.07	4.46*
Error	6.71	28	.24	
Within-patients	4.32	30		
Parental figure	.10	1	.10	<1.00
Premorbid $\times$ Parent	.19	1	.19	1.36
Error	4.03	28	.14	
Total	12.10	59		

\* Significant at the .025 point, one-tailed.

gued that the presence of a highly prepotent avoidance tendency vitiated discriminative response patterns in this group, and thus led to grossly overgeneralized avoidance behavior in the Poors. Garmezy (1952; Rodnick & Garmezy, 1957) has found evidence to support such a position when performance is observed under conditions of experimentally imposed social censure. In an attempt to evaluate the possibility that patterned avoidance was precluded as a result of the prepotence of avoidance behavior in the Poors in the present situation, matched subgroups of Goods and Poors were selected on the basis of total avoidance behavior. Nine individually matched pairs of Goods and Poors were found, each of which differed less than .2 point on total avoidance. Discriminative avoidance scores were computed for these 18 subjects by dividing avoidance scores on Card 6BM by the corresponding scores on 7BM. Thus, scores less than unity indicated greater avoidance in response to the father card, while scores greater than unity reflected more avoidance to the mother figure. The Mann-Whitney  $U$  test was applied to the groups, and found to be non-significant ( $U = 37.0$ ,  $p > .05$ ), although the rank difference was in the predicted direction. Accordingly, no evidence of consistent patterning of avoidance behavior was found, even when total avoidance was controlled. A further analysis of these matched subgroups did prove suggestive, however. The comparison of group variability of the discriminative avoidance scores indicated considerable heterogeneity between groups, with the Goods significantly more variable than the Poors ( $F = 3.87$ ,  $df = 8/8$ ,  $p < .05$ ). The ranges of these scores were .42–1.65 for the Goods, and .77–1.29 for the Poors. Since other studies have consistently found Goods to be more uniform than Poors in their performance on a number of tasks (Garmezy & Rodnick, 1959; Rodnick & Garmezy, 1957), this result is unexpected and suggests that discriminative avoidance behavior may have occurred in the Goods, although without the characteristic pattern predicted.

## DISCUSSION

The results of the present study demonstrate a significant difference in the pattern of

anxiety related behavior aroused in Good and Poor premorbid schizophrenics in response to pictorial representations of parental figures. The differential cue properties of these figures conform to expectations arising from the contrasting familial backgrounds to which the groups have presumably been exposed. Therefore, substantial support is found for the inferred asymmetry in parental dominance and submission in the backgrounds of these groups, with Goods coming from father dominant homes and Poors from a mother dominated background. It is also of interest to note the absence of group differences in total anxiety related behavior expressed. Under the present conditions, the groups cannot be distinguished on the basis of this over-all effect.

A reliable group difference in the level of avoidance related behavior was also found to characterize these patients. That is, under essentially comparable levels of total anxiety, Poors were significantly more avoidant in response to both parental figures. This result is in agreement with other observations pointing to the increased prepotence of avoidance behavior in patients attaining Poor premorbid adjustments, and thus lends further support to the position advanced by Rodnick and Garmezy (1957). The present data further demonstrate the utility of distinguishing patients along a continuum of premorbid personality adequacy. The efficacy of distinguishing subgroups of schizophrenics, apart from diagnostic subtype, has been debated for some years (Bellak, 1958; Garmezy & Rodnick, 1959). The present data concerning the patterning of anxiety and the differential pre-



potence of avoidance tendencies support the contention that such distinctions are warranted in terms of identifying subpatterns of response characteristics.

The results of the third hypothesis are less clear. In failing to confirm the predicted difference in the patterning of avoidance behavior, the data fail to agree with predictions drawn from the Rodnick-Garmezy position. Several limitations must be considered in qualifying the interpretation of this result, however. First, it is possible that the high level (prepotence) of avoidance behavior in the Poors interfered with discriminative avoidance responding, and hence concealed patterns which may be present under lower levels of response. Subgroups of Goods and Poors were selected to partially control for this possibility by equating total avoidance behavior. Comparison of discriminative avoidance scores in these matched groups also failed to reveal the differences predicted. Groups equated in this way necessarily match patients within an intermediate range of avoidance, however, and do not provide an unequivocal comparison. More adequate controls would be provided by patients on tranquilizing drugs, or patients who have achieved clinical remission of schizophrenic symptoms. A further difficulty is the order of presentation of the thematic cards. Direct evidence on card order is not available, although convention would lead one to expect that Card 6BM preceded 7BM in most cases. If recovery from threatening (censorious) stimuli is delayed in schizophrenics, the patterned (discriminated) response in Poors would be reduced, since this group is assumed to be more avoidant of the maternal figure, i.e., Card 6BM, given first. In view of these limitations in experimental controls, clear interpretation of the failure of the third hypothesis cannot be made from the present results. Evidence was found, however, to suggest that Goods are more discriminative in their avoidance behavior, since score variability for this group was significantly greater than for Poors. This result must be considered tentative, in view of the small sample size and the absence of subsidiary variables capable of reducing this variance to meaningful patterns.

## SUMMARY

Thematic productions in groups of Good and Poor premorbid schizophrenics were analyzed in terms of the relative amounts of anxiety and avoidance related imagery produced in response to cards representing parent-child relationships. Three hypotheses were specified concerning the patterning of these responses: Poor premorbid patients were expected to produce more anxiety related imagery in response to the mother than the father figure, while Goods were expected to show the reverse effect; Poors were expected to be more avoidant in response to both parental figures; and the pattern of avoidance behavior between groups was expected to differ.

The first two hypotheses were supported by the present data. These results were interpreted as supporting the position advanced by Rodnick and Garmezy and associates. The differential patterning of anxiety behavior in these groups is congruent with patterns of parental dominance and conflict assumed to have occurred in the developmental backgrounds of the patients. The increased level of avoidance behavior in the Poors is also in accord with the presumed increased prepotence of avoidance tendencies in this group. The prediction concerning differential patterning of avoidance behavior failed, however, to attain significance.

## REFERENCES

- BATESON, G., JACKSON, D., HALEY, J., & WEAKLAND, J. Toward a theory of schizophrenia. *Behav. Sci.* 1956, 1, 251-264.
- BELLAK, L. *Schizophrenia: A review of the syndrome*. New York: Logos, 1958.
- BINDRA, D. *Motivation: A systematic reinterpretation*. New York: Ronald, 1959.
- BOWEN, M. Family relationships in schizophrenia. In A. Auerback (Ed.), *Schizophrenia: An integrated approach*. New York: Ronald, 1959. Pp. 147-178.
- DOLLARD, J., & MOWRER, O. H. Measuring tension in written documents. *J. abnorm. soc. Psychol.* 1947, 42, 3-32.
- EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.
- FARINA, A. Patterns of role dominance and conflict in parents of schizophrenic patients. *J. abnorm. soc. Psychol.*, 1960, 61, 31-38.
- GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of



- reward and punishment. *J. Pers.*, 1952, 20, 253-276.
- GARMEZY, N., FARINA, A., & RODNICK, E. H. The structured situational test: A method for studying family interaction in schizophrenia. *Amer. J. Orthopsychiat.*, 1960, 30, 445-452.
- GARMEZY, N., & RODNICK, E. H. Premorbid adjustment and performance in schizophrenics: Implications for interpreting heterogeneity in schizophrenia. *J. nerv. ment. Dis.*, 1959, 129, 450-466.
- GARMEZY, N., STOCKNER, CAROL, & CLARKE, A. R. Child-rearing attitudes of mothers and fathers as reported by schizophrenic and normal control patients. *Amer. Psychologist*, 1959, 14, 333. (Abstract)
- LIDZ, T., & FLECK, S. Schizophrenia, human integration, and the role of the family. In D. Jackson (Ed.), *The etiology of schizophrenia*. New York: Basic Books, 1960. Pp. 323-345.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MAHL, G. F. Exploring emotional states by content analysis. In, *Trends in content analysis*. Urbana: Univer. Illinois Press, 1959. Pp. 89-130. (a)
- MAHL, G. F. Measuring the patient's anxiety during interviews from "expressive" aspects of speech. *Trans. N. Y. Acad. Sci.*, 1959, 21, 249-257. (b)
- MEADOW, A., GREENBLATT, M., & SOLOMON, H. C. *Frontal lobes and schizophrenia*. New York: Springfield, 1953.
- MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science*. New York: McGraw-Hill, 1959. Pp. 196-292.
- OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 348-423.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, 117, 515-525.
- PHILLIPS, L., & SMITH, J. G. *Rorschach interpretation: Advanced technique*. New York: Grune & Stratton, 1953.
- RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109-183.
- SCOTT, W. A. The avoidance of threatening material in imaginative behavior. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 572-585.
- WALKER, E. L., & ATKINSON, J. W. The expression of fear-related motivation in thematic apperception as a function of proximity to an atomic explosion. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. Pp. 143-159.
- WYNNE, L. C., RYCKOFF, I. M., DAY, JULIANA, & HIRSCH, S. J. Pseudo-mutuality in the family relations of schizophrenics. *Psychiatry*, 1958, 21, 205-220.

(Received April 10, 1961)

## RESPONSE TO CONFLICT AS DETERMINED BY PRIOR EXPOSURE TO CONFLICT<sup>1</sup>

LEONARD WORELL

Oklahoma State University

Clinical descriptions frequently contain statements indicating that people who are faced with strong conflicts demonstrate inefficient behavior not only in relation to these strong conflicts but also in resolving comparatively mild conflicting situations. A "neurotic" individual, for example, who presumably reacts with strong conflicting tendencies only in certain areas, may be characterized as generally indecisive. Thereby, the implication is conveyed that the person's exposure to severe conflict has significantly contributed to his inefficient handling of weaker conflicts. In spite of these recurrent observations, the problem of the generalization of the effects of conflict has received scant experimental attention. With small exception research in the area of conflict has concentrated on (a) the behavioral consequences of variations in several characteristics of conflicting responses, such as their relative strengths, approach-avoidance, etc., and (b) the role of individual differences in conflict resolution (Andreas, 1958; Berlyne, 1960; Block & Peterson, 1955; Brown, 1942; Cartwright, 1941; Castaneda & Worell, 1961; Hovland & Sears, 1938; Kaufman & Miller, 1949; Scodel, Ratoosh, & Minas, 1959; Sears & Hovland, 1941; Worell & Castaneda, 1961b; Worell & Hill, in press. This paper continues (Worell & Castaneda, 1961a) the investigation of a third area; that in which conflict serves as an antecedent to performance in other than the conflict situation. The general question under consideration is as follows: Does responding to conflict have an effect on the resolution of relatively weak conflict situations?

Our attention will be restricted to two separate but related facets of this general problem. First, we will describe two opposing

theoretical formulations that deal with the effects on the individual of exposure to different degrees of conflict. These hypothetical consequences will then be used to anticipate the performance of subjects in a subsequent conflict situation. Second, we will consider the issue of the degree of conflict itself. In this connection, the importance for performance of the absolute and relative strengths of conflicting tendencies will be examined.

The two theoretical approaches will be designated as the *competing response* and the *dynamogenic* hypotheses. These hypotheses may be, respectively, subsumed under the long-standing distinction made by psychologists between associative or learned and nonassociative approaches. According to the competing response view, it is assumed that responding to different levels of conflict leads to the learning of differing conflict-specific responses. For example, a person who is protractedly exposed to relatively strong conflict might be expected to learn such responses as withholding a decision or considering each alternative more carefully, etc. Then, in new but similar situations, the individual may be expected to invoke those behaviors which he has previously learned. Thus, with the competing response view, the effects of conflict are expected to have a limited generality—limited by the similarities between earlier and later conflict situations.

Sharply contrasted to the foregoing position is the dynamogenic view which proposes that conflict generates motivational consequences. The general suggestion that conflict may possess tension arousing or drive properties has been frequently made (Brown & Farber, 1951; Lewin, 1933; Miller, 1944; Miller & Stevenson, 1936). In addition, several studies designed to test this suggestion have agreed in finding heightened force of response as the degree of conflict is increased (Castaneda & Worell, 1961; Finger, 1941; Worell & Cas-

<sup>1</sup> The present research was assisted by a grant from the Research Foundation of Oklahoma State University.

taneda, 1961b). These results are consistent with the view that one consequence of conflict is an elevation of drive or motivation and that increasing degrees of conflict produce greater increments in drive. If conflict generates drive, then the effects of conflict on performance in a contiguously presented task would be expected to be evident irrespective of the similarity between the conflict and later performance situation. This is based on a view of drive as an indiscriminate energizer of performance (Hull, 1943).

Findings relevant to these alternatives were recently obtained by Worell and Castaneda (1961a), where the generalization of the effects of conflict was examined in relation to learning. In two experiments, subjects were exposed to either high or low experimentally induced conflict prior to the presentation of each pair of two verbal paired-associates tasks. Different degrees of conflict were aroused by the simultaneous exposure of two lines of either equal or unequal length, while the paired-associates lists contained words with either high or low amounts of intralist competition. The findings indicated that the conflict conditions did not significantly affect performance.

The foregoing findings are not compatible with expectations derived from the dynamogenic hypothesis. On the other hand, according to the competing response view, no necessary effects of conflict would be anticipated since the conflict and learning conditions were quite dissimilar. Consequently, in the present study, the degree of similarity between the conflict situation and the subsequent performance situation has been increased. Descriptively, subjects here are exposed to varying degrees of discrimination conflict and are then all placed into a weak discrimination conflict situation. Thus, the responses required in both the earlier and later conflict situations are the same, but the stimulus similarity changes to varying degrees for the subjects in the various conflict groups. In these situations, our interest is in a particular performance measure—namely, the time to initiate a decision or reaction evocation latency.

Given a weak conflict situation following differential conflict training, the following

predictions would be made with the competing response and dynamogenic positions. From the standpoint of the competing response view, subjects who have been formerly trained under strong conflict should be slower in the weak conflict situation than those who had experienced less severe conflict training conditions. In this it is assumed that prior exposure to strong conflict has led to the learning of conflict-specific responses which are transferred to the new but similar situation. On the other hand, from the dynamogenic view, subjects who were formerly exposed to strong conflict should be faster than subjects exposed to weaker conflicts. This is based on two considerations: first, that increased conflict is associated with increased drive, and secondly, that increased drive, as determined by the Taylor Manifest Anxiety scale, has been found to produce faster speeds in the same relatively weak conflict situation as used here by Worell and Castaneda (1961b).

The second concern of this paper is with the problem of the degree or amount of conflict. Writings on conflict have frequently made reference to the importance of either or both the relative and absolute strengths of conflicting tendencies in defining amount of conflict. In general, studies designed to determine the effects of these two characteristics on temporal performance have varied the strengths of the conflicting tendencies by variations in either training procedures or stimulus intensity. The justification for using training procedures seems apparent; the logic for employing variations in the intensity of two or more simultaneously presented stimuli seems to be based on the repeated finding that increasing response strengths tend to be associated with increasing intensities of *single* stimuli.

The results of studies aimed at testing the assumption that the absolute strengths of competing tendencies are influential in determining the amount of conflict have been either negative or at best equivocal. For example, Berlyne (1957) exposed subjects to either two equally bright or two equally dim stimuli and found that the weaker intensities were associated with significantly *longer* response times. Castaneda and Worell (1961) found a similar, but insignificant, trend for



increased speed of response to occur with increased intensities in an experiment which presented subjects with either two relatively bright or two relatively dim lights separated by equal log brightness differences. None of these results is in accord with expectations based on absolute strengths (Andreas, 1958; Bilodeau, 1950), since increased intensities should have produced longer response times. However, Andreas has obtained results agreeing with these expectations in his finding that response times were lengthened with increased amounts of training to the conflict stimuli.<sup>2</sup> In contrast, Bilodeau, also employing training procedures designed to increase the strengths of the competing tendencies, failed to find a significant decrease in movement time as the absolute strengths increased.

In contrast to the findings with absolute strengths, the results of studies directed toward the effects of the absolute *difference* in strength between or the relative strengths of conflicting tendencies on temporal performance have been with few exceptions positive. Thus, studies by Bilodeau (1950), Brown (1942), Cartwright (1941), Castaneda and Worell (1961), Korman (1960), Worell and Castaneda (1961b), and Worell and Hill (1962) have agreed in finding that speed of response increases as the competing tendencies become more unlike in strength. However, no differences in temporal performance as a function of variations in relative strength have been obtained by both Andreas (1958) and Broadhurst (1957).<sup>3</sup>

Aside from empirical examinations of the problem of the amount of conflict, at least two theoretical formulations have been advanced which specifically incorporate either or both the relative and absolute strengths of conflicting tendencies in their approaches. First, in a detailed attempt to integrate a construct of frustration within a Hullian

framework, Brown and Farber (1951) developed the position that frustration arose as a consequence of the magnitudes of *both* the absolute and relative strengths of competing tendencies. Moreover, their proposal suggests that somewhat greater importance should be given to the absolute strengths in producing frustration. In contrast, in a recent paper (Worell, 1961) where an attempt was made to account for the performance of differentially motivated subjects in learning and performance tasks involving competing response tendencies, a set of theoretical propositions based *solely* on the relative strengths of conflicting tendencies was discussed. In view of the rather general interest in the problem of the degree of conflict, the present study will examine the effects of the relative and absolute strengths of conflicting tendencies on both immediate performance and performance in a subsequently encountered conflict task.

## METHOD

*Design.* All subjects were initially exposed to a low conflict condition in a brightness discrimination apparatus for a predetermined series of trials. In part, this procedure was used to assess whether the subjects in the subsequently separated groups demonstrated initially equivalent performances. Following exposure to low conflict, subjects were placed into either high, intermediate, or low conflict conditions for an additional fixed number of trials. The simultaneous presentation of either two equally dim or two equally bright lights constituted the high conflict condition. Intermediate conflict was determined by the exposure of a relatively bright and dim light, while low conflict consisted of the appearance of a very bright and very dim light. In all, five groups were given differential conflict training. Two groups were exposed to high conflict, two groups received intermediate conflict, and one group was given low conflict. Comparisons between the high, intermediate, and low conflict conditions provide information relevant to the effects of the relative strengths of conflicting tendencies, while comparisons between the two groups in each of the high and intermediate conditions are relevant to the role of the absolute strengths. To assess the effects of these conditions, all subjects were finally placed in the original low conflict situation.

*Apparatus.* The apparatus is the same as that described elsewhere (Worell & Hill, 1962). Briefly, it consisted of two windows set into a black panel, a start platform attached to the base of the panel with a visual ready signal directly above, and a push button on both the left and right edges of the start platform. A Standard Electric timer was

<sup>2</sup> The studies performed by Andreas (1958) and Broadhurst (1957) used measures of temporal performance that were composites of reaction evocation time and what may be termed "movement" time. Since the meaning of a composite measure may be expected to vary considerably depending on what is involved in the movement performance, it is difficult to compare the results of studies using such a measure with those using reaction evocation latency.

<sup>3</sup> See Footnote 2.

activated with the illumination of the windows and was terminated when the subject removed his hand from the start platform. The lights in both windows were turned off by the subject pressing either push button. Reaction latency, or time from the illumination of the windows to removal of the hand from the start platform, was recorded to the nearest 1/100 second. These reaction time scores were converted to reciprocals ( $1/RT \times 1000$ ), so that all performances are presented in terms of speed of response.

The stimuli consisted of four different brightnesses, 2.8, 8.8, 90.9, and 291.4 foot candles, obtained by measurements at each window with a Macbeth illuminometer. The two brightest stimuli (90.9 and 291.4) and the two dimmest (2.8 and 8.8) were separated by differences of approximately .50 log units, while the two intermediate stimuli (8.8 and 90.9) were separated by an approximate 1.0 log unit difference.

**Procedure.** Each subject, participating individually, was instructed to push the button below either the "brighter" or "dimmer" light, depending upon the condition to which he was assigned. The subject was informed that the experiment was concerned with the accuracy of visual discrimination. The instructions did not contain any implication that speed was considered important.

All subjects were given a total of 64 discrimination trials. The first 16 of these constituted familiarization trials with low conflict; the intermediate

TABLE 1  
MEAN SPEEDS OF RESPONSE

Group <sup>a</sup>	I <sup>b</sup>	Preconflict performance	Conflict training	Post- minus preconflict performance
I (291.4 and 291.4)	B	19.14	7.86	-3.98
	D	14.37	6.34	.43
II (2.8 and 2.8)	B	18.27	9.87	12.96
	D	15.72	8.11	8.67
III (291.4 and 8.8)	B	15.02	18.71	29.27
	D	15.29	19.68	37.30
IV (90.9 and 2.8)	B	20.00	22.24	20.42
	D	16.68	18.74	13.43
V (291.4 and 2.8)	B	17.91	23.23	27.54
	D	16.07	20.96	23.16

<sup>a</sup> The numerical values given next to each numbered group reflect the intensity values of the conflict stimuli.

<sup>b</sup> The designation (B) and (D) represents the Bright and Dim instructional condition, respectively.

24 trials presented the experimentally induced conflict conditions; and the final 24 trials presented low conflict to all subjects again. In those conditions where two unequally bright stimuli were administered, the position of the correct stimulus was randomized such that the same stimulus did not appear more than three consecutive times in the same location.

**Subjects.** The subjects were 70 volunteer females from introductory psychology who were randomly assigned, 14 per group, to one of five groups. Within each group the brighter of the two stimuli was designated as the correct stimulus for half of the group and the dimmer for the remaining half.

## RESULTS

The mean speed scores ( $1/RT \times 1000$ ) for all subgroups during the three major phases of the experiment are presented in Table 1. Figure 1 contains the data for the five major conflict groups with the bright-dim conditions combined over successive eight trial blocks. The data for the various groups were analyzed and will be discussed separately for the three major phases of the experiment.

**Preconflict performance.** In order to determine whether the subjects in the conflict

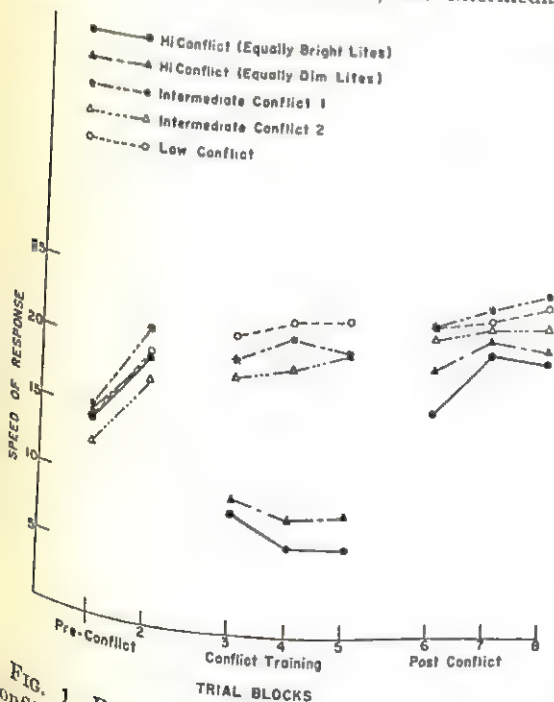


FIG. 1. Response speeds over trials for the five conflict groups during preconflict, conflict training, and postconflict conditions. (Each trial block represents the average performance on eight trials. Intermediate Conflict 1 corresponds to Group III; Intermediate Conflict 2 to Group IV [see Table 1].)



groups demonstrated comparable performances prior to exposure to conflict, a Lindquist (1953) Type III analysis of variance was performed on the preconflict training performances. The summary of this analysis appears in Table 2 and indicates that there were only two significant effects; that for the bright-dim instructional condition was significant at less than the .05 level, while the trials effect was significant well beyond the .001 level. Of particular interest is the absence of significant performance differences between the groups who were subsequently given the differential conflict treatments. The finding that speed increases with trials has been obtained many times before and indicates that performance becomes more rapid with practice. On the other hand, the finding of a significant difference between the bright-dim instructional conditions is of some interest. An examination of Table 1 indicates that there is a strong tendency for the subjects who were instructed to respond to the brighter stimulus to be faster. This result is in accord with a portion of the results obtained in a previous study (Castaneda & Worell, 1961). It would appear that through previous experience individuals may learn to respond to the "brighter" stimuli when relatively bright stimuli are present, and that forcing them to respond to dimmer stimuli may arouse interfering or incompatible tendencies. In other words, the instruction to respond to the dimmer stimulus seems to have increased the degree of conflict.

*Conflict training.* The mean speeds of performance for the five conflict groups are presented in the middle portion of Table 1. A second Lindquist (1953) Type III analysis was performed on the data during conflict training and is summarized in the middle portion of Table 2. This analysis provided two significant effects; the main effect of conflict was significant well beyond the .001 level as was the interaction of trials and the conflict conditions. An examination of the performances of the conflict groups during conflict training in Figure 1 indicates quite clearly the greater speeds of the low and intermediate conflict groups when compared with the two high conflict groups. In relation to the significant trials and conflict interaction, a further inspection of Figure 1 shows that while the low and intermediate conflict groups tended to increase in speed over trials, the high conflict groups tended to become slower. The foregoing results support the view that the conflict training conditions produced marked speed differences between the various groups.

A further analysis was made of the simple effects to determine whether the absolute strengths of competing tendencies provided significant differences in addition to those that are apparent for the relative strengths of these tendencies. It will be recalled that the two high conflict groups were exposed to either two equally bright or two equally dim stimuli. Since the difference in strength between the two stimuli was the same for either

TABLE 2  
SUMMARIES OF ANALYSES OF VARIANCE

Source	Preconflict performance			Conflict training			Postconflict performance		
	MS	df	F	MS	df	F	MS	df	F
Instructions (I)	13,353.64	1	4.73*	8,771.38	1	—	21.69	1	3.98**
Conflict (C)	2,306.62	4		130,938.53	4		7,620.97	4	
I × C	1,564.40	4	—	1,796.17	4	24.38***	441.82	4	—
Error (b)	2,825.01	60	271.29***	5,370.64	60	—	1,914.47	60	21.91***
Trials (T)	60,108.72	1		38.32	2		4,812.09	2	
T × I	33.71	1	—	30.57	2	—	322.41	2	—
T × C	174.32	4	—	933.33	8	5.69***	46.35	8	—
T × I × C	292.94	4	—	157.96	8		431.09	8	
Error (w)	221.57	60	—	164.15	120	—	219.58	120	—

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .



pair of stimuli, the major difference between the conditions is in terms of the absolute intensity of the stimuli. Likewise, one pair of intermediate conflict stimuli was made up of the 291.4 and 8.8 brightnesses, while the other pair consisted of the 90.9 and 2.8 brightnesses. In each case, there was about a 1.5 log unit difference between the stimuli. Assuming that within the present range of values, equal log differences represent subjectively equal differences, the major difference between the two intermediate conflict conditions also is in terms of the absolute intensity of the stimuli. Analyses revealed that the differences between the two high conflict groups and between the two intermediate conflict groups were not significant. Thus, there is no evidence for an effect of the absolute intensities on conflict performance.

*Postconflict performance.* The principal interest of this study was in the effects of prior exposure to different levels of conflict on performance in weak conflict situations. In order to determine these effects, an analysis was performed on the difference scores for each individual between preconflict performance (Trials 9-16) and postconflict performance (divided into successive eight trial blocks). The mean values for postconflict minus preconflict performance are presented in Table 1. The summary of an analysis of variance performed on these data may be seen in the third column of Table 2. This analysis provides two significant effects; the effect of trials was significant beyond the .001 level and the effects of conflict beyond the .01 level. Once again, the trials effect indicates that performances tend to become more rapid with practice (Figure 1). Of particular interest, however, is the significant conflict effect which suggests that previous exposure to various levels of conflict led to significant differences in the speed of resolving weak conflict situations. Reference to Table 1 and Figure 1 indicates that the two high conflict groups were slower than the two intermediate and lone low conflict groups. Moreover, the differences between these groups tend to persist over the entire 24 trials that (the weak conflict condition) were presented in Figure 1.

## DISCUSSION

The principal findings provide clear support for the view that the effects of conflict generalize to other related but less conflictful situations. It was found that individuals whose behavioral efficiency was impaired by exposure to relatively severe conflicts also demonstrated impaired ability to resolve subsequent weaker conflicting tendencies. When the present results are contrasted with those obtained in previous work (Worell & Castaneda, 1961a), there is a strong indication that generalization of the effects of conflict occurs along dimensions of similarity. Thus, in the former study where there was a minimal similarity between the conflict situation and the learning conditions, the generalization of conflict effects was only manifest when one took into consideration the nature of the individual's response to conflict. In the present study where there was a relatively high degree of similarity between the conflict situation and the subsequent performance task, a striking effect of conflict on later performance in a weaker conflict situation was found. From the standpoint of clinical approaches, these findings are in accord with what would be expected if one views personality as being a function of both the individual and the situation. Thus, the tendency to characterize an individual as generally indecisive may be adequate for a few persons, but that better prediction of behavior might be achieved by taking into account the specific nature of previously experienced conflicts and the present conflict situation.

The behavior of subjects in weak conflict following exposure to varying severities of conflict is consistent with expectations derived from the competing response formulation. It will be recalled that this hypothesis proposed that there would be a generalization of competing responses from the conflict training situation to the weak conflict situation, such that the behavior of subjects in weak conflict following performance under strong conflict would be significantly impaired. At the present time the exact nature of these hypothesized competing responses is unknown. Such responses, however, as hesitancy or holding back a response or learning to spend more time in looking from one alternative to the

other appear to be likely possibilities. Although the identification of these responses would be of some value, the consistency of the obtained data with the general competing response position is of most immediate importance.

Although the findings support the competing response formulation, they do not eliminate the importance of the dynamogenic view, or at least a modified version of this view. For example, it might be maintained that the greatest amount of emotionality or drive was aroused in both high conflict groups, but that this heightened emotionality acted in concert with the competing response tendencies. Thus, the tension aroused by the conflict conditions might be seen as having increased the general level of drive and thereby augmented the strengths of the competing tendencies. Such a "feedback" view has been discussed by both Brown and Farber (1951) and Worell (1961). Although admittedly speculative, this reasoning does serve to highlight our lack of independent evidence about the degree of emotionality engendered in any of our groups. In any event, whatever the status of conflict produced drive may be under the present circumstances, it seems apparent that a primary place must be made for some sort of associative (and in this case a competing response) formulation.

Apart from the competing response and dynamogenic views, an alternative explanation based on failure may be considered. It is possible that the various conflict training conditions were seen as differentially difficult by the subjects and that this varying difficulty produced differing levels of feelings of failure. In relation to this point, it should be noted that an attempt was made to minimize the possibility of failure occurring by having every response made by the subject to the conflict situation be a correct one. Thus, at all times the conflict situation was removed by the subject pressing the button under either simultaneously exposed light. Even if feelings of failure were not eliminated by this procedure, however, all that a precisely defined "failure" hypothesis would provide in terms of prediction would be a specification of the kinds of competing responses aroused as a function of conflict. In short, the general

predictions made in this experiment would not differ from those made with the competing response formulation, since the "failure" hypothesis would be a special case of the competing response formulation.

In addition to these considerations, the findings relevant to the role of the absolute and relative strengths of conflicting tendencies deserve some attention. Our results suggest that the strength or degree of conflict varies with the *difference* in strength between competing tendencies. Hence, the more closely competing tendencies approximate one another in strength the greater the degree of conflict. On the other hand, we have found that the absolute strengths do not contribute significantly to performance. Here, subjects exposed to either high conflict condition or to either intermediate conflict condition did not differ significantly in speed. As a result, within the present definition of absolute strength, strength of conflict appears to be principally, if not exclusively, based on the difference in strength between competing tendencies. Aside from the fact that other ways of manipulating absolute strength may provide different results, it is worth remembering that the present conflict conditions might be regarded as primarily involving conflict between two approach responses. Whether conflict situations which contain avoidance tendencies will provide similar results remains to be seen.

#### SUMMARY

As a partial outgrowth of the frequent clinical observation that individuals who are faced with strong conflicts show inefficient handling of these conflicts and weaker conflicts as well, the problem of the generalization of the effects of conflict was investigated. The first question that was examined was whether prior exposure to conflict affected performance in a new but similar conflict situation. In order to anticipate performance in a conflict situation following differential conflict training, two alternative hypotheses, termed the competing response and dynamogenic, were considered. According to the competing response view, exposure to conflict was assumed to lead to the learning of certain conflict-specific responses and performance in a new but similar conflict situation was ex-



pected to be affected by the generalization of these responses from previous conflict experiences. According to the dynamogenic view, conflict was assumed to generate drive and performance in a subsequent conflict situation would be affected as a function of augmented drive levels. The second interest of this study was in the problem of the degree or strength of conflict. An attempt was made to determine the role of the absolute and relative strengths of conflicting tendencies in performance.

To test these considerations, five groups of subjects were exposed to three successive phases of an experimental procedure. Initially, all subjects were exposed to a weak conflict situation. Following this, the subjects in the various groups were given training in responding to one of five levels (severities) of conflict. Finally, all subjects were again exposed to the original weak conflict situation. The findings indicated (a) that prior exposure to strong conflict led to a significant impairment of speed performance in weak conflict, and (b) that the degree of conflict was dependent on the relative strengths of or the difference in strength between competing tendencies and not on the absolute strengths of these tendencies. The significance of these results for our hypotheses and personality predictions was discussed.

## REFERENCES

- ANDREAS, B. B. Motor conflict behavior as a function of motivation and amount of training. *J. exp. Psychol.*, 1958, **55**, 173-178.
- BERLYNE, D. E. Conflict and choice time. *Brit. J. Psychol.*, 1957, **48**, 106-118.
- BERLYNE, D. E. *Conflict, arousal, and curiosity*. New York: McGraw-Hill, 1960.
- BILODEAU, INA MCD. Conflict behavior as a function of the strengths of competing response tendencies. Unpublished doctoral dissertation, State University of Iowa, 1950.
- BLOCK, J., & PETERSEN, P. Some personality correlates of confidence, caution, and speed in a decision situation. *J. abnorm. soc. Psychol.*, 1955, **51**, 34-41.
- BROADHURST, P. L. Emotionality and the Yerkes-Dodson law. *J. exp. Psychol.*, 1957, **54**, 345-352.
- BROWN, J. S. Factors determining conflict reactions in difficult discriminations. *J. exp. Psychol.*, 1942, **31**, 272-292.
- BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465-495.
- CARTWRIGHT, D. Relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174-193.
- CASTANEDA, A., & WORELL, L. Differential relation of latency and response vigor to stimulus similarity in brightness discrimination. *J. exp. Psychol.*, 1961, **61**, 309-314.
- FLINGER, F. W. Quantitative studies of "conflict": I. Variations in latency and strength of the rat's response in a discrimination-jumping situation. *J. comp. Psychol.*, 1941, **31**, 97-127.
- HOVLAND, C. I., & SEARS, R. R. Experiments on motor conflict. I. Types of conflict and their modes of resolution. *J. exp. Psychol.*, 1938, **23**, 477-503.
- HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
- KAUFMAN, E. L., & MILLER, N. E. Effect of number of reinforcements on strength of approach in an approach-avoidance conflict. *J. comp. physiol. Psychol.*, 1949, **42**, 63-74.
- KORMAN, M. Ego strength and conflict discrimination: An experimental construct validation of the ego strength scale. *J. consult. Psychol.*, 1960, **24**, 294-298.
- LEWIN, K. Environmental forces. In C. Murchison (Ed.), *Handbook of child psychology*. Worcester, Mass.: Clark Univer. Press, 1933. Ch. 14.
- LINDQUIST, E. R. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.
- MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald, 1944. Ch. 14.
- MILLER, N. E., & STEVENSON, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205-231.
- SCODEL, A., RATOOSH, P., & MINAS, J. S. Some personality correlates of decision making under conditions of risk. *Behav. Sci.*, 1959, **4**, 19-28.
- SEARS, R. R., & HOVLAND, C. I. Experiments on motor conflict. II. Determination of mode of resolution by comparative strengths of conflicting responses. *J. exp. Psychol.*, 1941, **28**, 280-286.
- WORELL, L. A theory of conflict: Implications for some motivational studies of performances. Unpublished manuscript, Oklahoma State University, 1961.
- WORELL, L., & CASTANEDA, A. Individual differences in conflict resolution and the ease of learning. *Amer. Psychologist*, 1961, **16**, 400. (Abstract) (a)
- WORELL, L., & CASTANEDA, A. Response to conflict as a function of response-defined anxiety. *J. Pers.*, 1961, **29**, 10-29. (b)
- WORELL, L., & HILL, L. Ego strength and anxiety in discrimination conflict resolution. *J. consult. Psychol.*, 1962, in press.

(Received April 10, 1961)



## CRITIQUE AND NOTES

### THE EFFECTS OF THREE TYPES OF AGGRESSION ON VASCULAR PROCESSES<sup>1</sup>

JACK E. HOKANSON AND MICHAEL BURGESS

*Florida State University*

Several recent investigations (Hokanson & Shetler, 1961; Worchel, 1957) have presented findings which suggest that postfrustration expression of aggression reduces general physiological arousal. Worchel, using verbal aggression, inferred this reduction in arousal from elevated postaggression scores on an intellectual task; whereas Hokanson and Shetler measured the course of systolic blood pressure over frustration and physical aggression. The purpose of the present study is to investigate more systematically the variable, type of aggression. To this end, systolic blood pressure and heart rate are measured before and after frustration and after each of four aggression conditions: Physical, Verbal, Fantasy, and No Aggression.

#### METHOD

Eighty (56 female and 24 male) white, college age subjects were used in a  $2 \times 4$  factorial design experiment with 3 males and 7 females in each cell. Half of the subjects were frustrated by the experimenter and subsequently placed in one of four aggression conditions: expressing Physical, Verbal, Fantasy, or No Aggression. The remaining 40 subjects were in a nonfrustrated control group and subsequently were also placed in one of the four aggression conditions. Subjects were assigned to conditions on a random basis with the restriction of maintaining equal male-female proportions in each cell. Heart rate was measured continuously and systolic blood pressure at about 2-minute intervals over the entire procedure, with the critical measurements being before and after frustration and immediately after opportunity to aggress.

Subjects were introduced to the experiment as one involving physiological responses to performance on intellectual tasks. To aid in establishing this deception, subjects were administered the Picture Completion subtest of the WAIS following an initial 8-minute adaptation period. During the second "task" of the session, the frustration conditions were introduced. The 40 subjects in the High Frustration condition were asked to count backwards from 100 to 0 by two's as quickly as they could. Following a standardized procedure during this counting, the experimenter repeatedly interrupted and harassed the

subject for counting too slowly, having the subject start over three times. After 1 minute the experimenter asked that the counting be stopped, stating that this data could not be used owing to the subject's uncooperative attitude. Subjects in the No Frustration condition were allowed to count backwards to zero without interruption, with the experimenter remarking "Good" at the completion of the series.

Immediately following the frustration procedures, the aggression phase of the experiment took place. The 20 subjects in the Physical Aggression condition were instructed as follows:

The next task involves an interpersonal guessing problem. You are to think of a number between one and ten and I [the experimenter] will try to guess the number. If I am wrong, you are to depress this plunger which will administer an electric shock to me. If I am correct, simply nod, and go on to the next trial. In this way I will be studying the effects of physical pain on guessing behavior.

The subject was given 10 opportunities to shock<sup>2</sup> the experimenter with subjects in both the High and Low Frustration conditions depressing the plunger an average of 8 times.

The 20 subjects in the No Aggression condition went through exactly the same procedure as described above with one exception: when they depressed the plunger, a small light was turned on which, they were told, signaled to the experimenter his error. Again, each group signaled errors an average of eight times.

The Verbal Aggression condition consisted of having the subject fill out a brief questionnaire evaluating the experimenter's capabilities as an experimenter, in the experimenter's presence. A typical item was to rate the experimenter's competency in relating to subjects.

Finally, 20 subjects were asked to create a story to Card 8BM of the TAT. This card was selected for the Fantasy Aggression condition owing to its moderate picture-pull towards aggressive themes (Eron, 1950).

Each aggression condition took 2 minutes, following which systolic blood pressure and heart rate were again noted. Blood pressure was measured via a sphygmomanometer operated by the experimenter.

<sup>1</sup> This project was supported by the Research Council of Florida State University.

<sup>2</sup> Although the experimenter was wired to an elaborate "shock" apparatus, he did not actually receive a shock, but merely behaved as if he had.

The heart rate apparatus consisted of a rubber diaphragm transmitting pulses from the carotid artery through a pneumatic system, with a pen marking pulses on a constant speed kymograph.

## RESULTS

Analyses of the changes in systolic blood pressure and heart rate following the frustration manipulations indicate that the High Frustration condition has significantly higher elevations on both measures. High Frustration subjects manifest a mean increase of 10.3 millimeters of Hg and 9.0 beats per minute, while subjects in the Low Frustration condition increase on the average only 3.2 millimeters and .75 beats per minute, respectively. Both of these differences are significant at the .01 level.

Table 1 presents the analysis of variance and results of the Duncan multiple range test for change in systolic blood pressure after the aggression opportunities, with prefrustration base level as reference.

The significant main effect for Frustration indicates that relative to subjects in the Low Frustration condition, frustrated subjects manifest a significantly more elevated systolic blood pressure immediately after the aggression phase of the experiment (5.8 vs. 2.9 millimeters of Hg above prefrustration base level).

The significant  $F$  ratios for Aggression and the Frustration  $\times$  Aggression interaction may best be interpreted by reference to the cell means and

TABLE 1

ELEVATION OF SYSTOLIC BLOOD PRESSURE AFTER AGGRESSION, RE PREFRUSTRATION BASE LEVEL

Source	df	SS	MS	F
Frustration	1	163	163.0	10.00*
Aggression	3	311	103.7	6.36*
Frustration $\times$ Aggression	3	434	144.7	8.88*
Within	72	1176	16.3	
Total	79	2084		

## Cell Means of Elevation in Systolic Pressure

Condition	Aggression			
	Physical	Verbal	Fantasy	None
Low Frustration	3.2 <sub>a</sub>	3.4 <sub>a</sub>	2.2 <sub>a</sub>	2.8 <sub>a</sub>
High Frustration	1.2 <sub>a</sub>	1.8 <sub>a</sub>	9.2 <sub>b</sub>	10.8 <sub>b</sub>

Note.—Cells with different subscripts are significantly different from one another at the .05 level, by Duncan multiple range test.  
\*  $p = .01$ .

TABLE 2

ELEVATION OF HEART RATE AFTER AGGRESSION, RE PREFRUSTRATION BASE LEVEL

Source	df	SS	MS	F
Frustration	1	352.8	352.8	9.28*
Aggression	3	505.0	168.3	4.43*
Frustration $\times$ Aggression	3	1038.2	346.1	9.11*
Within	72	2736.0	38.0	
Total	79	4632.0		

## Cell Means of Elevation in Heart Rate

Condition	Aggression			
	Physical	Verbal	Fantasy	None
Low Frustration	-1.2 <sub>a</sub>	2.8 <sub>ab</sub>	-1.8 <sub>a</sub>	-0.2 <sub>a</sub>
High Frustration	0.8 <sub>a</sub>	-3.8 <sub>a</sub>	8.0 <sub>bc</sub>	11.4 <sub>c</sub>

Note.—Cells with different subscripts are significantly different from one another at the .05 level, by Duncan multiple range test.  
\*  $p = .01$ .

range test results at the bottom of Table 1. The systolic elevations, re base level, of subjects under High Frustration-Physical Aggression (1.2) and High Frustration-Verbal Aggression (1.8) are not significantly different from the elevations in their respective Low Frustration control cells (3.2 and 3.4). On the other hand, the systolic elevations of subjects under the High Frustration-Fantasy Aggression (9.2) and High Frustration-No Aggression (10.8) are significantly greater ( $p = .05$ ) than the elevations in their control conditions (2.2 and 2.8) and in the High Frustration-Physical and Verbal Aggression cells.

Reference to Table 2 and the cell means at the bottom of the table suggests substantially the same pattern of results with the heart rate measure. Frustrated subjects who were given a chance to physically or verbally aggress manifested returns in heart rate to +.8 and -3.8 beats per minute with respect to base level. These values are not significantly different from the rates in the Low Frustration control conditions (-1.2 and +2.8 beats per minute, respectively). Cell means under High Frustration-Fantasy Aggression (8.0) and High Frustration-No Aggression (11.4) are again significantly ( $p = .05$ ) higher than the means in either their respective Low Frustration control conditions (-1.8 and -.2 beats per minute) or the High Frustration-Physical and Verbal Aggression cells (-.8 and -3.8 beats per minute, respectively).

To summarize the vascular changes: frustrated subjects manifest a mean increase of 10.3 milli-



meters of Hg and 9.0 beats per minute as a result of the frustration manipulation; frustrated subjects given an opportunity to physically or verbally aggress against the frustrator show a return of systolic pressure and heart rate to levels not significantly different from their Low Frustration control levels; whereas subjects given an opportunity to aggress in fantasy, or given no opportunity, maintain significantly elevated pressures.

### DISCUSSION

The results of this study offer support for the hypothesis that the expression of aggression towards a frustrator reduces frustration produced physiological arousal. Several considerations, however, tend to limit the generality of this hypothesis. Type of aggression apparently is an important variable owing to the failure of the fantasy aggression condition to lower the elevated vascular processes. While the time allowed for fantasy aggression was relatively short (2 minutes) in this study, it may be said that within this period opportunity for the direct communication of physical or verbal aggression to the frustrator is substantially more efficient in reducing systolic pressure and heart rate.

Contrary to attempts by previous authors (Feshbach, 1955; Worchel, 1957), the present investigators caution against concluding that the results reflect a reduction in "aggressive drive." Indeed, the current research was instituted in an effort to overcome the methodological difficulties encountered when attempting to demonstrate such drive diminution. Feshbach's finding that "residual aggression" towards a frustrator (measured by Sentence Completion test and questionnaire) is reduced by interpolated opportunity for fantasy aggression has been questioned by Berkowitz (1958). Owing to the possible arousal of guilt and/or aggression anxiety on the part of his fantasy aggression subjects, Feshbach's results could be explained on the basis of response suppression, rather than drive reduction. Certainly, the failure of the current investigation to find any arousal reduction with fantasy aggression raises further questions about Feshbach's conclusions.

Worchel (1957), on the other hand, assumes a reduction in "aggressive drive" on the basis of heightened intellectual performance with his "opportunity to (verbally) aggress" group. While tenable, this conclusion is at least one step away from his data; and a more parsimonious explanation would be that performance is increased due to general "tension" reduction. This latter con-

clusion is consistent with the present data involving verbal aggression.

With respect to the current study it should be noted that both the motor activity involved in expressing aggression and the time allowed for expressing aggression can be ruled out as factors affecting the course of the vascular processes. In the Physical Aggression and No Aggression conditions, the physical activity demanded of the subject was the same, the only difference being that in the former (arousal reducing) condition the subject was aware that he was physically hurting the experimenter. Along the same lines, in the verbal and fantasy aggression conditions, the subject engaged in approximately the same amount of writing, with again only the former condition involving the direct communication of aggression to the experimenter. In addition, since the time allowed for each type of aggression was 2 minutes, any effects due to homeostatic reduction of vascular processes would be the same for each condition. It would appear then that the critical factor in the arousal reduction was the subject's knowledge that he was hurting the frustrating experimenter either physically or verbally.

A final point of interest is a comparison of the amount of aggression expressed between frustrated and nonfrustrated subjects for each aggression condition. Under physical aggression, both number of shocks administered and average pressure exerted on the plunger were measured. (Both these measures were found to be valid indicators of increased aggression in a previous study by Hokanson, 1961.) In the Verbal Aggression condition, the degree of negative evaluation of the experimenter was measured by a three-item Likert-type questionnaire. Fantasy aggression was measured via a TAT hostility scale previously successfully used by Hokanson and Gordon (1958). For each type of aggression, the difference in amount or intensity of aggression expressed between frustrated and nonfrustrated subjects was not significant. This rather unexpected finding raises a question of whether the degree of aggression expressed is related to the "cathartic" effects found. A similar question was posed recently by Rosenbaum and deCharms (1960) in a study which found that frustrated subjects expressed only a meager degree of aggression towards the frustrator but still exhibited reduced residual aggression. While the current data do not allow any unequivocal conclusions on this point, perhaps the critical feature in this paradigm is merely *having the opportunity to aggress* towards the frustrator following a frustration.



## SUMMARY

Eighty college age subjects underwent either a frustrating or nonfrustrating experience and subsequently were placed in one of four aggression conditions: Physical, Verbal, Fantasy, or No Aggression. The intent of this  $2 \times 4$  factorial design experiment was to measure the effects of each type of aggression on frustration produced elevations in systolic blood pressure and heart rate.

The results indicated that frustrated subjects who were given an opportunity to physically or verbally aggress against the frustrator, manifested returns on both physiological measures to the levels of the Low Frustration control subjects. Frustrated subjects in the Fantasy and No Aggression conditions exhibited significantly elevated systolic pressures and heart rates at the postaggression recording of these measures.

The results were discussed as offering support for the hypothesis that the expression of Physical or Verbal aggression towards a frustrator tends to reduce general, physical arousal.

## REFERENCES

- BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.*, 1958, **55**, 257-283.
- ERON, L. D. A normative study of the thematic apperception test. *Psychol. Monogr.*, 1950, **64**(9, Whole No. 315).
- FESTIBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3-11.
- HOKANSON, J. E. The effects of frustration and anxiety on overt aggression. *J. abnorm. soc. Psychol.*, 1961, **62**, 346-351.
- HOKANSON, J. E., & GORDON, J. E. The expression and inhibition of hostility in imaginative and overt behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 327-333.
- HOKANSON, J. E., & SHETLER, S. The effect of overt aggression on physiological arousal. *J. abnorm. soc. Psychol.*, 1961, **63**, 446-448.
- ROSENBAUM, M. E., & DECHARMS, R. Direct and vicarious reduction of hostility. *J. abnorm. soc. Psychol.*, 1960, **60**, 105-111.
- WORCHEL, P. Catharsis and the relief of hostility. *J. abnorm. soc. Psychol.*, 1957, **55**, 238-243.

(Received March 20, 1961)

## THE EFFECT OF AFFECTIVE TONE ON THE VERBAL BEHAVIOR OF PROCESS AND REACTIVE SCHIZOPHRENICS<sup>1</sup>

ALAN S. DEWOLFE<sup>2</sup>

*Northwestern University*

Increasing attention is currently being paid to the classification of schizophrenics into the so-called Process and Reactive groups. The former term usually denotes a poor premorbid adjustment and the latter a relatively adequate premorbid adjustment.<sup>3</sup> Unlike groups classified by the Kraepelinian system, Reactive and Process groups have been found to differ both prognostically, Reactives having the more favorable prognosis (Kantor, Wallner, & Winder, 1953; Phillips, 1953; Rodnick & Garnezy, 1957; and others), and behaviorally (Becker, 1956; Bleke, 1955; Harris, 1957; among others).

Purely behavioral studies have yielded superficially contradictory results with the Reactives sometimes more efficient, accurate, etc., than the Process group and sometimes vice versa. Physiological studies, however, have yielded more consistent results. Using physiological measures, King (1958) using metholyl, a sympathetic nervous system depressant, and DeVault (1955) using situational stress, emotional pictures, found the Reactives showed a more intense reaction than the Process group who responded minimally, if at all. More recently, Zuckerman and Grosz (1959) have reported results contradictory to King, but studies

of schizophrenics classified as Remitter or Non-remitter, a distinction expected to be correlated to a relatively high degree with the Process-Reactive classification, by Funkenstein, Greenblatt, and Solomon (1952) and Hirschstein (1955) found results similar to King's and thus provide indirect support for his findings. These physiological studies suggested the generalization that the Reactive schizophrenics' response to affective stimulation is greater than that of the Process' response and is usually equal to or greater than the Normals' response.

The physiological findings further suggest that the previously mentioned inconsistencies in behavioral level comparisons of Process and Reactive groups may be the result of differences in reactivity. Attempts by the present writer to classify tasks and/or situations as stressful or affective led to the hypothesis that Reactives tend to be superior in neutral situations (e.g., Becker, 1956; Bleke, 1955; Harris, 1957) but inferior in affective situations. Studies by Williams and Machi (1957) and Zubin, Windle, and Hamwi (1953) using the Remitter vs. Non-remitter classification found results to support both the Reactives' superior performance on neutral tasks, perceptual-motor tasks and rote learning of nonmeaningful material, and inferior performance in situations with affective connotations (poorer performance on intellectual and memory tasks and slower reaction times in word association tests).

The less efficient performance of Reactives in stress or affective situations may be due to interference phenomena, emotional responsiveness leading in some way to response tendencies which interfere with performance, as is sometimes the case with normal subjects. The Reactives being more responsive, like Normals, may, therefore, exhibit more interference effects than the less responsive Process group. Tasks unfortunately are not always clearcut on affectivity-non-affectivity and their results, therefore, provide only suggestive evidence for the hypothesis outlined above.

The present experiment was designed to investigate with a purely behavioral task, the hypothesis that Reactive schizophrenics are more responsive to affective stimuli than Process schizophrenics and that this emotional respon-

<sup>1</sup> This article is based on a doctoral dissertation submitted to the Graduate School of Northwestern University. The author wishes to express his sincere appreciation to Janet A. Taylor for her aid, advice, and encouragement during the investigation. He is also indebted to the departments of psychology and psychiatry at the Downey Veterans Administration Hospital, the Hines Veterans Administration Hospital, and the Chicago West Side Veterans Administration Hospital, for their cooperation in making subjects and facilities available.

<sup>2</sup> Now at the Psychology Service, Veterans Administration Hospital, Hines, Illinois.

<sup>3</sup> Investigators have not uniformly agreed as to whether there are two classes of schizophrenia or whether the groups are polarities of a continuum. Lack of clearcut bimodal distribution in any of the measuring systems makes either view possible. However, regardless of the investigator's predilections, research has typically involved selection and investigation of the two extreme groups. Usage of the terms Process and Reactive in the present study is meant only to reflect this procedure and not necessarily advocacy of the position of separate types of schizophrenia.

siveness may in certain types of tasks lead to response interference. The behavioral response measures consisted of the personal pronoun "I" as a measure of personal reference and reaction time in making up sentences. The personal reference measure was intended to show the degree of emotional responsiveness as judged by an avoidance reaction, i.e., a decrease in personal reference under the threat of the affectively toned verbs, while the reaction time measure was invoked to demonstrate that emotional reactivity may lead to response interference and "poorer" performance, i.e., increased reaction time. According to speculations offered above, the greater reactivity of the Reactives was expected to result in the greatest *decrease* in rate of the use of "I" (avoidance response) and the greatest *increase* in reaction time (interference effect) with the affective verbs in the Reactive group. The smallest differences were anticipated in the Process group and an intermediate magnitude of difference in the Normal group on both measures. The order of the effect of affective stimuli was expected to be the same in both measures with the direction of change opposite in the two measures, i.e., a decrease in personal reference and an increase in reaction time. The greatest effect was anticipated in the Reactives, the least in the Process group, with the Normals showing a magnitude of effect between two schizophrenic groups on both measures.

#### METHOD

**Subjects.**<sup>4</sup> Two schizophrenic groups and one Normal Control group, each consisting of 15 male inpatients from Veterans Administration hospitals were used. All subjects, schizophrenic or normal, were between the ages of 20 and 55 with a minimum of an eighth grade education, a current hospitalization of under 9 months duration, and by history had neither inherited nor acquired organic brain pathology.

The schizophrenic subjects met these additional criteria: no more than two nonmilitary psychiatric hospitalizations including the current hospitalization and no form of convulsive or coma therapy for 9 months prior to the experiment. Schizophrenic subjects all had an established diagnosis of schizophrenic reaction and were classified as Process or Reactive by scores on the Phillips (1953) Scale of Premorbid Adjustment in Schizophrenia. The Process group scored 18 or over on the scale and the Re-

actives 14 or under. The Normal Control subjects were selected from the orthopedic and tuberculosis services and had not been hospitalized more than 9 months in the 2 years prior to the current hospitalization and had no history past or present of significant abnormal emotional response.

In the three groups, the subjects were individually matched for age, educational level, and socioeconomic level. Educational level was scaled on three intervals: not a high school graduate, high school graduate, high school graduate with more than one semester of college. Socioeconomic status was scaled by occupational level in a three-level modification of the Centers (1949) index.<sup>5</sup> The subjects from each major group were assigned in counterbalanced order to one of two subgroups which differed only in which neutral task was presented first. All subjects who successfully made up the 150 sentences were used. Any sentence using one of the pronouns presented was accepted regardless of the accuracy of perception of the stimulus verb or the grammatical correctness or meaningfulness of the sentence.

**Materials and Apparatus.** Stimuli for the experimental task were 150 slides. Each slide contained a different past tense verb in the center and four pronouns (HE, SHE, THEY, and I) on a single line below the verb. All words were in upper-case letters. The slides were projected on a screen by a 35-millimeter projector. The 150 slides were divided into three groups of 50, two groups with verbs neutral in meaning and one with affective verbs. The affective and neutral verbs were selected by six judges—advanced graduate students in psychology or experienced clinical psychologists.

The order of pronouns on the slides was varied systematically and the Affective and Neutral slide groups were equated for order of pronouns. The order of presentation of the slides within the three groups, affective and two neutral, was randomized by shuffling slides within the groups between-subjects. To measure reaction times, the projector was equipped with a single switch which simultaneously activated a clock and projected the slide on the screen. The experimenter stopped the clock with a separate hand switch when the subject responded.

**Procedure.** Each subject was tested individually in a single period of about an hour. When the subject entered the room and was seated, the experimenter moved to the projector behind the subject and darkened the room sufficiently to make the experimental stimuli readily visible, but did not totally darken the room. Subjects were instructed that they would see slides with a centered verb and four pronouns, that their task was to make up sentences, and that any kind of a sentence was acceptable as long as they used one of the pronouns and the verb. These instructions were projected on the screen. Then, any

<sup>5</sup> A more detailed description of the matching and other experimental procedures, as well as more inclusive theoretical discussions can be found in the original dissertation available from University Microfilms, Inc., Ann Arbor, Michigan (L. C. Card No. Mic 6-06542).

<sup>4</sup> Details on subjects and their selection have been deposited with the American Documentation Institute. Order Document No. 7090 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.



TABLE 1  
ANALYSES OF VARIANCE OF "I" FREQUENCY DATA

Source	A. All conditions analysis			B. Neutral 1-Neutral 2 analysis		
	df	MS	F	df	MS	F
Groups	2	362.8	1.27	2	324.1	1.52
Between-subjects same group	42	285.2		42	213.7	
Total between-subjects	44			44		
Conditions	2	207.6	6.68*	1	9.4	0.47
Groups $\times$ Conditions	4	42.9	1.38	2	61.8	3.12**
Pooled Subjects $\times$ Conditions	84	51.1		42	19.8	
Total within-subjects	90			45		

\*  $p < .01$ .

\*\*  $.05 < p < .10$ .

misunderstanding about the instructions and any task relevant questions were answered.

The first slide was presented and if any aid was required, the experimenter repeated the instructions. Each subject was given the 150 slides: 50 neutral, then 50 affective, and finally 50 more neutral with the particular set of neutral slides given first counter-

balanced within each group.<sup>a</sup> At regular intervals in each set of 50 slides, subjects were encouraged by a general statement like, "That's fine, you're doing well." After every sentence, the subjects were reinforced by the word "umhmm" spoken by the experimenter regardless of the pronoun used, i.e., the task was used to study the effect of affective tone on performance rather than the conditioning process. The subjects were prepared for the next stimulus by the experimenter saying, "All right, ready," prior to each new slide.

## RESULTS

*Personal Reference.* The personal reference data were evaluated by an analysis of variance with Groups (Reactive, Process, and Normal) and Conditions (Neutral 1, Affective, and Neutral 2) as the main effects. Since all subjects were received all three conditions, this variable was treated as the correlated term in an analysis of repeated measures in all analyses in this study. The  $F$  of 6.68 for Conditions was significant ( $p < .01$ ) with all others not significant (Table 1). An analysis of variance on the initial neutral condition showed that the groups were far from

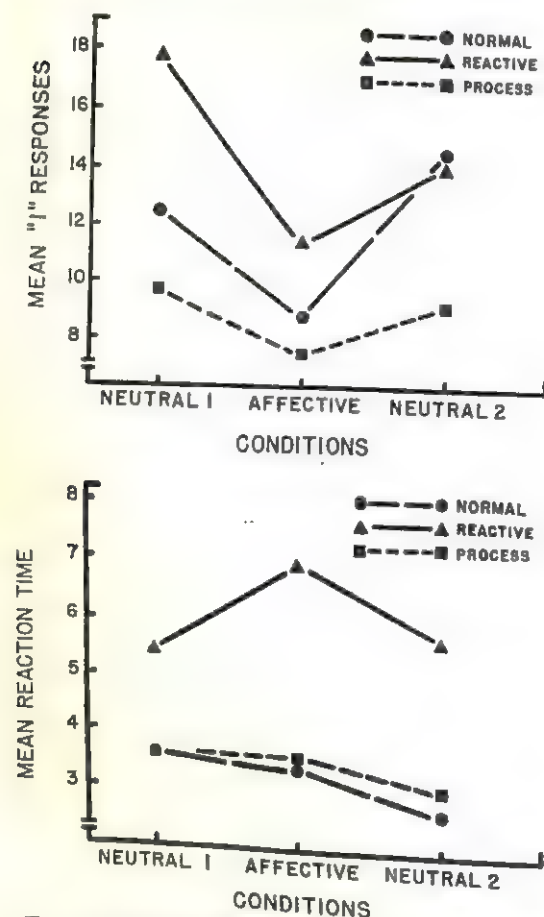


FIG. 1. Number of personal references per 50 trials and mean reaction time in seconds per 50 trials in the Affective and Neutral conditions.

<sup>a</sup> Subjects were presented with a group of neutral verbs first and affective second, and never the reverse. This procedure was followed since findings of the Rodnick and Garmezy (1957) suggest that if the affectively toned verbs were presented first, the effects of this material might persist during the Neutral condition and depress measures of the difference between the two conditions. Further, the groups might be differentially affected with the Reactives expected to be most responsive and slowest to dissipate their emotional response showing the greatest effect. In order to test for the presence of perseveration effects, as well as to provide a check on the possibility that differences in the Affective condition were due primarily to practice effects, a second group of neutral verbs was added following the affective slides.

significantly different in rate of personal reference in this condition ( $F < 1.00$ ).

In the analysis of the two neutral conditions (Table 1), the  $F$  for Groups was not significant, but the Groups  $\times$  Conditions interaction neared significance ( $.05 < p < .10$ ) which was probably the result of the Normals showing an increase in personal reference between the two neutral conditions (upper portion of Figure 1). In this analysis, the Conditions  $F$  was far from significant indicating that differences in the analysis with the Affective condition present were not the result of time or trials. In the analyses of the "I" data, Bartlett tests showed heterogeneity of variance significant between the .01 and .05 levels which was not considered sufficiently extreme to warrant rejection of the conclusions from these analyses.

These analyses verified the hypothesis of the general effect of the affective tone variable. Although the Groups  $\times$  Conditions interaction  $F$  was not significant, some indirect support for the hypothesis of the differential effect of the affective stimuli was found in the fact that the groups differed in the expected direction and fell in the predicted order of magnitude of difference between conditions. Analysis by direct difference  $t$  tests within the groups showed that the Reactives ( $t = 2.74$ ) and Normals ( $t = 3.08$ ) had highly significant ( $p < .01$ ) differences between the neutral and affective conditions, while the Process group ( $t = 1.31$ ) did not differ significantly.

**Reaction Time.** As indicated in the lower portion of Figure 1, the Reactives displayed a sharp increase in mean reaction time in the Affective condition and their mean reaction time in the final Neutral condition was slightly higher than in the initial Neutral condition. The Process and Normal groups showed negligible change in the Affective condition, followed by a slight drop in

mean reaction time in the final Neutral condition.

In the initial analysis of variance with reaction time data (Table 2), both main effects and the Groups  $\times$  Conditions interaction effects were significant. In  $t$  test analyses, the group did not differ significantly in the initial Neutral condition. In the Affective and final Neutral conditions, the Reactives had a reaction time significantly higher than the other groups which did not differ from each other. Comparing the two neutral conditions (Table 2), the Groups  $F$  was significant ( $p < .05$ ). The Conditions and Groups  $\times$  Conditions interaction neared significance ( $.05 < p < .10$ ) apparently due primarily to the magnitude of the drop in mean reaction time between the first and final Neutral conditions in the Process and Normal groups (lower portion of Figure 1).

Bartlett tests on the reaction time data uniformly showed extreme heterogeneity of variance (all  $B$ s having a  $p < .001$ ). Therefore, the reaction time data was reanalyzed by distribution free analyses. The Groups variable was evaluated by a Kruskal-Wallis test and Conditions by a Cochran  $Q$  test. Neither was significant. As a distribution free analysis equivalent of the Groups  $\times$  Conditions interaction effect, the groups were compared by a Kruskal-Wallis test using the difference between the first neutral and affective conditions as the raw scores. The  $H$  of 7.10 was significant ( $p < .05$ ). The two kinds of analyses then, gave parallel significant results on the predicted differences in reaction time among the groups.

**Additional Analyses.** The relationship between the two measures—reaction time and frequency of using "I"—was studied by means of Spearman rho's. Using the totals for both measures, the rho for all subjects was  $-.36$  ( $p < .05$ ). In analyzing this relationship in the groups taken

TABLE 2  
ANALYSES OF VARIANCE OF MEAN REACTION TIME DATA

Source	A. All conditions analysis			B. Neutral 1-Neutral 2 analysis		
	<i>df</i>	<i>MS</i>	<i>F</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Groups	2	115.0	4.39**	2	58.6	3.78**
Between-subjects same group	42	26.2		42	15.5	
Total between-subjects	44			44		
Conditions	2	7.6	6.90*	1	4.5	3.46***
Groups $\times$ Conditions	4	3.5	3.18**	2	3.2	2.46
Pooled Subjects $\times$ Conditions	84	1.1		42	1.3	
Total within-subjects	90			45		

\*  $p < .01$ .

\*\*  $p < .05$ .

\*\*\*  $.05 < p < .10$ .



separately, the rho's were as follows: Reactive — .60 ( $p < .05$ ), Normal — .65 ( $p < .05$ ), and Process .13 ( $ns$ ).

### DISCUSSION

The hypothesis that Reactive schizophrenics are more emotionally responsive than Process schizophrenics was confirmed. The evidence further suggests that the Process schizophrenics are less responsive than Normals. These results using affective verbs with behavioral measures parallel the results of DeVault (1955) who used emotional pictures and a physiological measure, indicating that the greater emotional responsiveness of Reactive schizophrenics appears at both levels—the behavioral and physiological.

The Reactives' significant increase in reaction time in the Affective condition coupled with the Process group's slight decrease supported the speculation that the inferior performance of the Reactives on some laboratory tasks may be due to some interference phenomenon. That the interference phenomenon is associated with differences in emotional responsiveness is supported within this study by the not statistically significant but parallel results (same order of responsiveness) in the personal reference measure and by the significant negative correlation between the two measures. The direction of changes was predicted to be opposite in these measures. The association between interference effects and responsiveness is further supported by the parallel results between the reaction time measure and DeVault's (1955) physiological measure of emotional responsiveness. As indicated earlier, the interference phenomenon hypothesis would account for differences found between Process and Reactive groups by Becker (1956), Bleke (1955), Harris (1957) and others and the lack of these differences under different conditions by Rodnick and Garnezy (1957), as well as explaining the seemingly contradictory results within the studies of Williams and Machi (1957) and Zubin, Windle, and Hamwi (1953) with Remitters and Nonremitters.

It was noted that only the Reactives showed the interference effect (increase in reaction time) with the affective verbs (lower portion of Figure 1). The fact that the Reactives were also higher in reaction time (not significantly, however) in the initial neutral condition potentially raises two conflicting explanations: the hypothesis that change is proportional to the original level of reaction time, or the hypothesis that a certain original level of responsiveness is necessary for interference to occur, or threshold hypothesis.

Although the results of this study were not directly applicable to this problem and, therefore, are inconclusive, there was limited support for the latter hypothesis. When the reaction time in the original Neutral condition was correlated with the change in reaction time between initial Neutral and Affective conditions within the three groups, a significant relationship appeared only in the Reactive group. The Reactive group rho of .55 was significant at the .05 level, while the Process group rho of .05 and the Normal rho of — .35 were not significant. This supports the threshold hypothesis which, assuming the relationship between responsiveness and reaction time, would predict this relationship in the Reactives, while the proportional model would predict this relationship in all groups. Combining the above and the suggestive personal reference data, the present writer tentatively suggests the possibility of relationships between interference phenomena, emotional responsiveness, and what might be called "personalization" of the experimental situation, "ego involvement," or some similar construct. These speculations, however, must await substantiation from further research.

### SUMMARY

The present investigation was designed to study emotional responsiveness in Process and Reactive schizophrenics. Process and Reactive groups, classified by level of premorbid adjustment on the Phillips scale, were compared to each other and a Normal Control group on a task eliciting verbal behavior. Each group consisted of 15 hospitalized male subjects. The subjects were required to make up sentences using a given verb and any one of four pronouns. They were presented with a series of 50 verbs neutral in meaning, followed by 50 affective verbs, and finally, a second series of 50 neutral verbs.

The comparisons of major interest involved differences in reaction time and personal reference (use of "I") measures between neutral and affective conditions. It was hypothesized that the order of groups, when compared on the magnitude of differences between these conditions, from greatest to smallest would be Reactive, Normal, and Process and that these differences reflect differences in responsiveness to affective stimuli in both measures and response interference in the reaction time measure. The same order of magnitude of change was predicted for both the measures, although an increase in reaction times and a decrease in personal reference was expected in the affective condition.

In the main, the hypothesis of differential re-



sponsiveness to affective stimuli was verified. In the personal reference measure, the differences between conditions were in the expected direction with the magnitude of differences in the expected order.

Confirmation of the hypotheses of differential emotional responsiveness and interference phenomena was found in the reaction time analyses. As expected, the Reactive group exhibited a significant increase in reaction time in the Affective condition which perseverated into the final Neutral condition, while the changes in the Normal and Process groups were not significant.

## REFERENCES

- BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, 53, 229-236.
- BLEKE, R. C. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *J. Pers.*, 1955, 23, 479-498.
- CENTERS, R. *The psychology of social class*. Princeton: Princeton Univer. Press, 1949.
- DEVULT, S. H. Physiological responsiveness in reactive and process schizophrenia. Unpublished doctoral dissertation, Michigan State University, 1955.
- FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. An autonomic nervous system test of prognostic significance in relation to electroshock treatment. *Psychosom. Med.*, 1952, 14, 347-362.
- HARRIS, J. G., JR. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *J. Pers.*, 1957, 25, 651-671.
- HIRSCHSTEIN, R. The significance of characteristic autonomic nervous system responses in the adjustment, change and outcome in schizophrenia. *J. nerv. ment. Dis.*, 1955, 122, 254-262.
- KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, 17, 157-162.
- KING, G. F. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *J. abnorm. soc. Psychol.*, 1958, 56, 160-164.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, 117, 515-525.
- RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109-184.
- WILLIAMS, R. J., & MACHI, V. S. An analysis of interpersonal correlations among thirty psychotics. *J. abnorm. soc. Psychol.*, 1957, 55, 50-57.
- ZUBIN, J., WINDLE, C., & HAMWI, VIOLET. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, 21, 342-355.
- ZUCKERMAN, M., & GROSZ, H. J. Contradictory results using the Mecholy Test to differentiate process and reactive schizophrenia. *J. abnorm. soc. Psychol.*, 1959, 59, 145-146.

(Received April 24, 1961)

## NUMBER OF VERBAL INTERRELATIONSHIPS AS A DETERMINANT OF GROUP SIZE

GEORGE F. CASTORE

*Mental Health Institute, Mt. Pleasant, Iowa*

Group size has been determined by common sense estimates, preferences of the group psychologists, or by traditional attitudes. This brief note is to present one consideration, number of verbal interrelationships as related to group size. The degree to which group size sets the limit on the possible interrelationships between members during a single session is likely to be related to the amount of individual participation which may be achieved.

Attention is given to the direction of remarks to others rather than the source of a response alone. The value for an individual is assumed to reside in an increase in communication potentials by increasing the range of persons to whom a given person can relate. An individual who extends his relationships to a wider range of persons is compelled to do more than just talk about himself or broadcast in a general way. To continue an established verbal relationship, he is involved in the complexity of understanding what another says, seeing how he is understood by another, dealing with specific individual differences in meanings and interpretations, observing how he influences others and then noting effects they have on him. To increase the range of persons with whom one person communicates is likely to increase the skills of communication such as clarifying, anticipating distortion, etc., and provide a broader cultural context within which he may improve and meet his social needs. The advantage to the group as a whole of increasing the amount of interpersonal participation is that a more varied and, hence, richer supply of verbal interaction arises with more opportunity for each person to find a relationship which meets his particular need of the moment. The greater variation of interpersonal interchange is likely to have the additional educational effects of broadening each person's interests and range of social sensitivity. To the psychologist the advantage lies in possession of a simple and reliable tool of observation.

The aim of a therapeutic group is the improvement of the patient. Conditions which increase rate, depth, or reliability of growth that is deemed desirable would generally be accepted as of value. One of these conditions would be op-

portunity for verbal expression. Opportunity is related to many facets of group interaction. For example, opportunity has a different significance when it is easy to obtain and when it is difficult. The same condition that makes opportunity easy for all is likely to provide a situation in which two members may hold the floor to the exclusion of others. The emergence of subgroups makes verbal expression more difficult for the individual members of a group. Groups which are too small may find that practices such as establishing of norms, forming subgroups, shunting responsibility to the group leader, and holding to stereotyped patterns of interaction, act along with other behaviors to reduce over-all interaction between members. Large groups possess as many obvious objections.

### METHOD

To determine the size of a small interaction group that would have optimal growth value for its members, a measure of the extent to which the members initiate verbal relationships with each other in varied size groups was used. This measure was not the frequency of participation. Two persons may produce high frequency of response. The measure was more one of diffusion of interaction, the number of other members to which each person directed at least one remark during an hour's session. In a group composed of eight members, perfect diffusion would be obtained if each member directed at least one remark to seven other members. For any size group, the number of possible interchanges would be represented by  $n(n-1)$ . A percentage index measure for each group was simply obtained by the formula

$$\frac{f_0}{n(n-1)}, \text{ and } \frac{\sum f_0}{\sum n(n-1)}$$

for each mean interval value when combining odd-even size groups.

A simple procedure was used to make the observations. A list was made of the members, sitting in a circle, beginning at the psychologist's left. Each person is given a successive number. When person Number 3 speaks to person Number 5, a 5 is written after the name of Person 3. No more record need be made of Person 3 responding to Person 5 during this meeting. Any other person to whom Number



3 speaks is indicated by writing in the appropriate number when the response occurs. The same procedure is used for recording the responses of each person present, excluding the psychologist. At the end of an hour's session, the record will show the extent of verbal relationships each person has made with the other members in the group.

To be recorded as a directed response, the response should possess one of the following four qualities: mention of the recipient's name; mention of a second person singular "you" in such a way that there is no doubt about the referent; a complete sentence answer to another person's question which was directed to the subsequent speaker; and looking directly at the person with whom he is speaking. The implication here is that only one person among those present can be identified as capable of meeting the needs of the speaker at that moment.

There are also some qualities which definitely exclude remarks from being recorded as directed ones. These are simple Yes-No answers or answers which are incomplete or incoherent sentences. A response which elaborates on the response of another but is not directed to anyone particularly as well as a broadcasting response which is said to the group as a whole or for anyone who chooses to be concerned, likewise is not credited. Where there is the slightest doubt as to whether or not a response is directed to a specific person, the response should not be tabulated. The type of response that may be described as "thinking aloud" most often is undirected.

Groups observed were inpatient therapeutic groups in mental hospital settings. The groups were well-practiced, the members of each group having met together for over a month. Consequently, the results given in this study represent a favorable condition of participation. It would be expected that members of groups in earlier stages of development would be less inclined to direct remarks to so many different persons present.

TABLE 1

COMPARING THE PERCENTAGE OF MEMBER INTER-RELATIONSHIPS AND AVERAGE NUMBER OF INTERRELATIONSHIPS FOR 55 OBSERVATIONS OF VARIED SIZE GROUPS

Group size	Number of groups observed	Average number interrelationships per group	Percentage of relationships possible $\frac{\sum f_o}{\sum n(n-1)}$
5-6	3	20.0	.800
7-8	11	37.1	.732
9-10	3	33.6	.440
11-12	8	57.8	.348
13-14	11	53.4	.318
15-16	8	57.8	.248
17-18	7	40.5	.141
19-20	4	50.0	.134

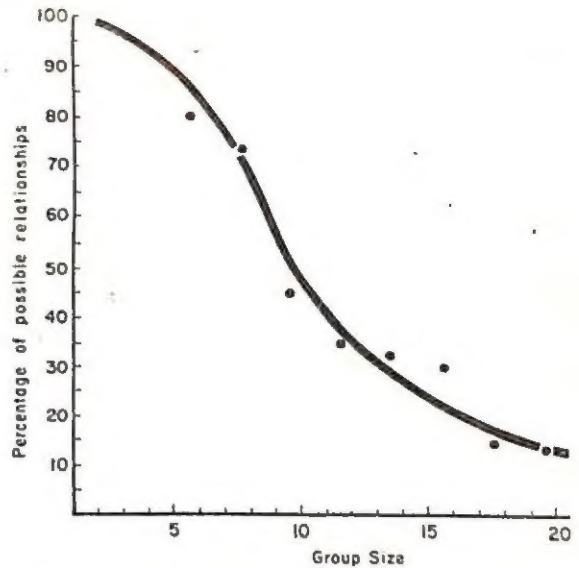


FIG. 1. Mean percentage values representing the extent of relationship initiated during verbal interchange between members of varied sizes.

The age range of the members varied from 14 to 70, the average representing the usual membership of psychotherapeutic groups in state institutions for treatment of severe personality disorders. No records were made on distribution of sex but, since heterogeneity was a feature of group composition, it is roughly estimated that males and females were equally distributed. All types of diagnostic categories were represented in these groups. There was little preponderance of any one type. Although most patients were considered as potentially active treatment patients, a few in each group were regressed. In all, 55 groups were observed with most of the observation falling between groups composed of 11 and 18 members.

## RESULTS

There is a marked reduction in the percentage measure, but not the observed number of interactions when group size reaches 9 members. This is shown in Table 1 by comparing average number of interrelationships per group with the percentage for each group size represented. There is also a marked change with 17 or more members present. Based on these data a desirable level of verbal interaction would be determined by the aims of a particular type group. Obviously, for those who are quite reticent and interact with extreme difficulty, more opportunity should arise in a group with less than 9 members. Others could profit from groups having up to 16 members.



Figure 1 shows the distribution of the observed number of initiated relationships translated into percentage of relationships possible. The suggested curve of distribution emphasizes more clearly the effect of group size on the extent of interrelationships that may be expected.

It is recognized that the types of interactions

may be a much more important consideration than their frequency of occurrence. Where the increase in participation between members is a desirable feature of group process, the findings presented are likely to be of value in group organization.

(Received April 10, 1961)

*Journal of Abnormal and Social Psychology*  
1962, Vol. 64, No. 6, 458-460

## RELATIONSHIP BETWEEN PROBABILITY OF ITEM ENDORSEMENT AND SOCIAL DESIRABILITY SCALE VALUE FOR HIGH AND LOW GROUPS ON EDWARDS' *SD* SCALE<sup>1</sup>

ALLEN L. EDWARDS AND JERALD N. WALKER

*University of Washington*

There is considerable evidence to show that the probability of a True response to a personality item can be predicted quite accurately from a knowledge of the social desirability scale value of the item (Cowen & Tongas, 1959; Edwards, 1953, 1957a, 1957b, 1959; Hanley, 1956). If the items in an inventory are scaled for social desirability, then a socially desirable or socially undesirable response to an item can be defined in terms of the scale value of the item. A socially desirable response is defined as a True response to an item with a socially desirable scale value or a False response to an item with a socially undesirable scale value. A socially undesirable response is defined as a True response to an item with a socially undesirable scale value or a False response to an item with a socially desirable scale value (Edwards, 1957b).

Edwards (1957b) has developed the Social Desirability (*SD*) scale to measure the tendency of subjects to give socially desirable responses to personality items. All of the 39 items in the *SD* scale are keyed for socially desirable responses, but 30 items are also keyed False and only 9 are keyed True. Because of the imbalance in True-False keying of the *SD* scale, Wiggins (1959), Jackson and Messick (1958), and Couch and Keniston (1961) have suggested that the *SD* scale may be measuring an acquiescent tendency rather than a tendency to give socially desirable responses. Evidence has been presented by Ed-

wards (1957b) and Edwards and Walker (1961) which indicates that acquiescence is not an important component of the *SD* scale.

The present study was designed to provide further evidence on the acquiescence and social desirability interpretations of scores on the *SD* scale. If low scores on the *SD* scale reflect the tendency to respond True and high scores the tendency to respond False, then such tendencies should operate over the entire range of the social desirability continuum. Thus, for High and Low *SD* groups, the regression lines of probability of a True response (*Y*) on the social desirability scale values (*X*) of a given set of items should, according to the acquiescence interpretation of the *SD* scale, be parallel with the regression line for the Low *SD* group having a higher *Y* intercept than that for the High *SD* group. Thus, the Low *SD* group should also have a higher mean probability of a True response than the High *SD* group.

A social desirability interpretation of the *SD* scale would result in the same prediction as the acquiescence interpretation with respect to the difference in the *Y* intercepts. However, the social desirability interpretation of the *SD* scale would predict that the regression lines would not be parallel but rather that they would cross, with the line for the High *SD* group having the steeper slope. The basis for this prediction is that the High *SD* subjects should have a greater probability of a True response than Low *SD* subjects for items with socially desirable scale values. For items with socially undesirable scale values,

<sup>1</sup> This research was supported in part by Research Grant M-4075 from the National Institute of Mental Health, United States Public Health Service.



TABLE 1

MEANS AND RANGES OF SCORES ON THE *SD* SCALE  
FOR THREE GROUPS OF SUBJECTS

Group	Male <i>N</i>	Female <i>N</i>	Total <i>N</i>	<i>X</i>	Range
High	42	43	85	36.588	35-39
Middle	22	48	70	32.686	31-34
Low	36	46	82	25.951	12-30

the difference in the probabilities should be reversed with the Low *SD* group having the greater probability of a True response. Furthermore, if scores on the *SD* scale are relatively uninfluenced by the tendency to respond True, then the mean probability of a True response for the High and Low *SD* groups should be much the same.

#### METHOD

The *SD* scale and 150 items from the EPPS were administered under standard instructions to 237 students at the University of Washington. Subjects were divided into three groups on the basis of their *SD* scores. Records from both sexes were pooled, since there is evidence (Edwards, 1957b) to show that little distortion is to be expected with such pooling. Table 1 gives the composition of each of the groups with their respective means and ranges on the *SD* scale.

Social desirability scale values of the 150 EPPS items were available from a previous study (Edwards, 1953). For each *SD* group, the probability of endorsement for each of the 150 EPPS items was obtained and these probabilities were correlated with the social desirability scale values of the items.

#### RESULTS AND DISCUSSION

Table 2 gives the means and variances of the probability of endorsement for the 150 EPPS items, for each of the separate groups. The means are somewhat higher than those obtained in another study (Edwards, 1959) in which the items of the Interpersonal Check List (ICL) were used.<sup>2</sup> The difference between the means in the two studies can be attributed to the differences in the social desirability scale values of the two sets of items. The set of EPPS items has a greater proportion of items at the desirable end of the social desirability continuum, than does the set of ICL items. Thus, we would expect to obtain a higher mean probability of endorsement for the EPPS items than for the items of the ICL.

According to the acquiescence interpretation of the *SD* scale, the Low *SD* group should have a higher mean probability of endorsement of the

<sup>2</sup> For the ICL items, the mean probability of endorsement for males was .48 and for females .46.

EPPS items than the High *SD* group, whereas the actual difference is in the reverse direction. In accordance with the social desirability interpretation of the *SD* scale, the difference between the means for the Low and High *SD* groups is not statistically significant ( $p > .70$ ).

The correlations between probability of endorsement and social desirability scale values of the items are .844, .710, and .731 for the High, Middle, and Low *SD* groups, respectively. The difference between the correlations for the High and Low *SD* groups is statistically significant ( $p < .01$ ).

It is clear, from the values of the regression coefficients for the separate groups, that the regression lines are not parallel but rather that they cross in accordance with the social desirability interpretation of the *SD* scale. As expected, the slope of the line for the High *SD* group is steeper than that for the Low *SD* group. The difference in the slopes of the regression lines is statistically significant ( $p < .01$ ).

#### SUMMARY

Predictions as to the nature of the regression lines of probability of endorsement on social desirability scale values for High, Middle, and Low *SD* groups were based on an acquiescence and a social desirability interpretation of the *SD* scale. The *SD* scale and a set of 150 items from the EPPS were administered to 237 college students. The subjects were divided into three separate groups on the basis of their scores on the *SD* scale. For each group separately, the regression line of probability of endorsement on social desirability scale values of the EPPS items was obtained. The form of the regression lines was that predicted by the social desirability interpretation of the *SD* scale, rather than that predicted by the acquiescence interpretation. Furthermore, contrary to the acquiescence interpretation, the mean probability of a True response to the EPPS items was not greater for the Low *SD* group than for the High *SD* group.

TABLE 2

MEANS, VARIANCES, AND REGRESSION COEFFICIENTS  
OF ITEM PROBABILITIES ( $\bar{Y}$ ) AND CORRELATIONS  
BETWEEN ITEM PROBABILITIES AND SOCIAL  
DESIRABILITY SCALE VALUES FOR  
THREE *SD* GROUPS

Group	$\bar{Y}$	$s_y^2$	$r_{xy}$	$b_y$	$s_{y.x^2}$
High <i>SD</i>	.645	.086	.844	.28	.025
Middle <i>SD</i>	.626	.087	.710	.24	.043
Low <i>SD</i>	.633	.069	.731	.22	.032

## REFERENCES

- COUCH, A., & KENISTON, K. Agreeing response set and social desirability. *J. abnorm. soc. Psychol.*, 1961, 62, 175-179.
- COWEN, E. L., & TONGAS, P. The social desirability of trait descriptive terms: Applications to a self-concept inventory. *J. consult. Psychol.*, 1959, 23, 361-365.
- EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, 37, 90-93.
- EDWARDS, A. L. Social desirability and probability of endorsement of items in the Interpersonal Check List. *J. abnorm. soc. Psychol.*, 1957, 55, 394-396. (a)
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957. (b)
- EDWARDS, A. L. Social desirability and the description of others. *J. abnorm. soc. Psychol.*, 1959, 59, 434-436.
- EDWARDS, A. L., & WALKER, J. N. Social desirability and agreement response set. *J. abnorm. soc. Psychol.*, 1961, 62, 180-183.
- HANLEY, C. Social desirability and responses to items from three MMPI scales: *D*, *Sc*, and *K*. *J. appl. Psychol.*, 1956, 40, 324-328.
- JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, 55, 243-252.
- WIGGINS, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *J. consult. Psychol.*, 1959, 23, 419-427.

(Received April 27, 1961)